26738

National Library of Canada | Bibliothèque nationale du Canada

CANADIAN THESES ON MICROFICHE | THÈSES CANADIENNES SUR MICROFICHE

NAME OF AUTHOR/*NOM DE L'AUTEUR* Phyllis Margaret Craig

TITLE OF THESIS/*TITRE DE LA THÈSE* Establishing Validity in Field Evaluation of Nurse Practitioners

UNIVERSITY/*UNIVERSITÉ* University of Alberta

DEGREE FOR WHICH THESIS WAS PRESENTED/ *GRADE POUR LEQUEL CETTE THÈSE FUT PRÉSENTÉE* Master of Health Services Administration

YEAR THIS DEGREE CONFERRED/*ANNÉE D'OBTENTION DE CE GRADE* 1975

NAME OF SUPERVISOR/*NOM DU DIRECTEUR DE THÈSE* Dr. Clarke B. Hazlett

Permission is hereby granted to the NATIONAL LIBRARY OF CANADA to microfilm this thesis and to lend or sell copies of the film.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

L'autorisation est, par la présente, accordée à la BIBLIOTHÈQUE NATIONALE DU CANADA de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

L'auteur se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans l'autorisation écrite de l'auteur.

DATED/*DATÉ* Oct 16/75 SIGNED/*SIGNÉ* Phyllis M. Craig

PERMANENT ADDRESS/*RÉSIDENCE FIXE* #1009 10853 - Saskatchewan Drive, Edmonton, Alberta.

NL-91 (3-74)

THE UNIVERSITY OF ALBERTA

ESTABLISHING VALIDITY IN FIELD EVALUATION OF NURSE PRACTITIONERS

by

PHYLLIS M. CRAIG

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF MASTER OF HEALTH SERVICES ADMINISTRATION

DIVISION OF HEALTH SERVICES ADMINISTRATION

DEPARTMENT OF COMMUNITY MEDICINE

EDMONTON, ALBERTA

FALL, 1975

© PHYLLIS M. CRAIG 1976

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled "Establishing Validity in Field Evaluation of Nurse Practitioners" submitted by Phyllis M. Craig in partial fulfillment of the requirements for the degree of Master of Health Services Administration.

.....
Supervisor

.....

.....

Date 8 October, 1975

ABSTRACT

Several Canadian universities provide Clinical Training for Nurses (CTN) courses for those working in an extended role in isolated Medical Services Branch nursing stations. While each university is responsible for evaluation of students upon course completion, no field evaluation has empirically established whether or not the CTN courses do influence the practice of the graduates. A project was undertaken to develop (1) instruments of known reliability and validity for the evaluation of representative skills and knowledge of CTN graduates, and (2) a sampling design which would allow assessment of whether or not the competencies of recent and non-recent graduates are equivalent, and equal to those of nurses without the CTN course.

The instruments developed included written examinations, and observation instruments for the assessment of skills in physical examination, history taking, and suturing. The sample for validation studies was a class of eight students, whose pre- and post-tests on the observation instruments were evaluated by a panel of judges including physicians, nurses, nurse practitioners, and a CTN nurse-educator. The several professional groups were utilized as judges to allow the estimation of the inter-judge reliability and inter-professional validity of the instruments. Patient simulations were employed to increase standardization of tasks, and audiovisual recordings made of the physicals and histories to enable independence in marking.

A two-tailed alpha of .05 was used, except for analyses of pre-post comparisons and correlations (one-tailed alpha of .05 used). Validation

procedures included face, content, concurrent and construct validation. The estimates of concurrent validity were the correlations among the judges for each observation instrument, and between the written examinations. The test for construct validity for each instrument was of mean differences (increases) from pre- to post-testing. Results indicated not only that all instruments had acceptable levels of both reliability and validity, but could be dependably used by any one of the professional groups which participated as judges in the study. An independent analysis of variance indicated that the instruments had sufficient construct validity to determine whether or not the behavior of CTN graduates differs from that of nurses without the course in the education setting.

Pre- and post-test data from a sample of 32 previous students were analyzed for antecedent variates hypothesized to affect performance. Two variates, previous length of experience in nursing stations and midwifery training were found to have influence, and thus the use of these variates as covariates in field data analysis was recommended. The design recommended involves random sampling of experimental and control groups stratified by the size of the community served, with the experimental group also stratified into recent and non-recent CTN graduates.

The limitations of this study included the small sample size for validation studies, as well as several assumptions which were made regarding the representativeness of the sample and judges. The recommendations arising from the project include those for utilization of the instruments and design in field evaluation, as well as the continued integration of reliability, validity, and theories from education into the conceptual framework of health care delivery.

ACKNOWLEDGEMENTS

In acknowledging those who have contributed not only to this project, but to my graduate studies, I shall name individually only two persons. Firstly, my gratitude is expressed to Dr. Clarke Hazlett for his continued guidance and support. I should also like to convey my appreciation to Dr. Shirley Stinson, who has encouraged and inspired my professional development over an extended period.

In the hope that each will recognize that their contribution has been sincerely appreciated, I should like to thank those friends and colleagues who were judges, patients, subjects and validators for this project.

I am also obliged to the University of Alberta's Nurse Practitioner Programme and Health Sciences Audiovisual Education, and to Medical Services Branch for their co-operation and assistance.

Fellowships from the Canadian Nurses' Foundation and the Canadian Red Cross Society which have facilitated my studies have been gratefully received.

Lastly, but with deepest appreciation, I wish to acknowledge the ongoing understanding, support and encouragement which I have received from my family and friends.

TABLE OF CONTENTS

	Page
ABSTRACT	iv
ACKNOWLEDGEMENTS	vi
TABLE OF CONTENTS	vii
LIST OF TABLES	xi
LIST OF FIGURES	xii
CHAPTER	
I. INTRODUCTION	1
Background to the Problem	1
Statement and Importance of the Problem	3
Description of the Study	4
Assumptions and Limitations of the Study	5
Definition of Terms	6
Format of the Thesis	6
II. REVIEW OF SELECTED LITERATURE	7
Evaluation of Nurse Practitioners	7
Introduction	7
Historical Antecedents	8
Indirect Evaluations	9
Education Programmes' Performance Evaluation	11
Performance Evaluation in Practice	13
Summary	16
Reliability of Measurement Instruments	17
Validity of Measurement Instruments	18
Introduction	18

Page

CHAPTER

Face Validity 20

Content Validity. 21

Concurrent and Predictive Validity. 22

Construct Validity. 23

Summary 25

Design Validity 26

Introduction. 26

Internal Validity 26

External Validity 27

Summary 29

Summary of Chapter. 30

III. METHODOLOGY 32

Introduction. 32

Validity in Measurements. 32

Criteria in Selection of Objectives 32

Instruments Developed 36

Validity and Reliability of Instruments 38

Written examinations. 38

Observation instruments 39

Validity in Design. 43

The Design Selected 43

Internal Validity 43

External Validity 45

Summary 46

Page

CHAPTER

IV. RESULTS AND INTERPRETATION OF DATA ANALYSES 47

Introduction. 47

Measurement Instruments 47

Areas of Objectives Selected. 47

Written Examinations. 49

Observation Instruments 51

Incorporation of Instruments into Field Design. 59

Sampling. 64

Summary 67

V. SUMMARY AND RECOMMENDATIONS 70

Summary 70

Recommendations 72

Recommendations for Field Evaluation. 73

Instruments 73

Sampling. 73

Evaluator 74

Length and time of visits 76

Staff preparation 77

Data analysis 77

Follow-up 79

General Recommendations 79

REFERENCES. 82

Page

APPENDICES

APPENDIX A: Written General and Paediatrics Examination. 88

APPENDIX B: Written Obstetrics and the Newborn Examination. 100

APPENDIX C: Physical Examination Instrument. 108

APPENDIX D: Adult History Instrument 112

APPENDIX E: Paediatric History Instrument. 115

APPENDIX F: Suturing Instrument. 118

APPENDIX G: Instruction Manual 120

APPENDIX H: Nursing Stations Stratified by Community Size 139

LIST OF TABLES

Page

TABLE

1. Stratification of Antecedent Variates Hypothesized to Influence Performance of RN's and/or CTN's 44

2. Content Areas Selected for Evaluation, by Criteria. 48

3. Correlations Among Written Examinations 49

4. Analysis of Pre-Post (RN-CTN) Mean Differences on Written Examination 50

5. Correlations of Judges on Observation Examinations: Inter-judge Reliability and Inter-professional Validity Estimates. 52

6. Mean Inter-judge Reliability and Inter-professional Validity Estimates for Observation Instruments. 54

7. Validity and Reliability Estimates of Physical Examination Instrument: Live and Video-taped Observations. 56

8. Central Tendencies and Dispersions of Raw Scores, and Reliability Estimates: Observation Instruments 58

9. F-values of Pre-Post (RN-CTN) Mean Differences in Raw and Standardized Scores: Observation Instruments . . . 60

10. T-tests (Welch's Prime Adjustments) of Mean Differences on Observation Instruments Using Pre- and Post-tests as Measures of Independent Groups 61

11. Influence of Antecedent Variates: Probabilities Associated with F-values under Null Hypothesis. 63

12. Retention Rates and Central Tendencies of Length of Service (in months) of CTN's Sponsored by Northern Region, MSB 68

LIST OF FIGURES

Page

FIGURE

1. Illustration of methodology of development of evaluation package 33

2. Applications of reliability coefficients. 55

3. Total patient visits in 1973 to all N.W.T. and five southern nursing stations, by month. 65

4. Sampling design for field evaluation. 66

CHAPTER I

INTRODUCTION

Background to the Problem

Since the mid-1960's, there has been a trend toward the introduction of paramedical personnel as physician extenders in primary care. Three factors central to this development have been a shortage of physicians, maldistribution and increasing specialization of physicians leading to inaccessibility of primary care in some areas, and spiralling costs of health care. In Canada, it has been agreed that nurses are best suited to act as physician extenders in primary care, but require additional educational preparation for the extended role (Boudreau, 1972).

The Canadian federal government does employ nurses in the primary care extended role. Health and Welfare Canada, through Medical Services Branch, provides health services to all native people, and to the Northwest Territories (N.W.T.) and Yukon. Although medical consultation and visitations are available for isolated communities, physicians are rarely resident in them. The basic providers of health care, both prevention and treatment, are registered nurses stationed in nursing stations in the remote areas. Some of these nurses have taken a federally sponsored training programme to improve their capabilities in this extended role.

Six courses were established for the upgrading of clinical skills of nursing station nurses following the recommendations of the

Kergin Report (1970), a federally commissioned study "to investigate the possibilities of providing a formal training program for nurses employed by the Medical Services Branch" (p. 1). These Clinical Training for Nurses courses (CTN courses) at the Universities of McGill, Sherbrooke, Western Ontario, Toronto, Manitoba, and Alberta are identified by different names at the various universities, but all were established to prepare nurses for their responsibilities in the extended role, particularly in nursing stations.

Under the auspices of Medical Services Branch, representatives of the six universities met in June, 1972 to discuss curriculum development, their progress, and problems. It was evident that student evaluation, a requirement of all programmes, was a common concern. An anticipated difficulty was field assessment of the graduates, as no specific performance requirements or detailed course content had been developed. In subsequent discussions, the universities decided to cooperatively build a bank of behavioral objectives (Mager, 1962) which would be used as the core curriculum for all CTN courses. The bank that was developed (some ten thousand behavioral objectives) was subjected to rigorous content validation by physicians and nurses. The results of this major undertaking, described elsewhere by Hazlett (1975), provided the specification of skills and knowledge necessary for practicing CTN graduates. Accordingly, this purposeful delineation of the professional competencies expected in this group of nurses in an extended role could be used as the criteria for any field evaluation.

Statement and Importance of the Problem

Although each university had assessed students upon course completion, nurses' competencies in this extended role prior to course influence were unknown. Further, it could not be assumed, nor was it ever demonstrated, that CTN graduates retained either their knowledge or enhanced clinical skill after returning to the nursing stations. A design was therefore needed which would determine whether or not a difference in practice existed between (1) those nurses who had completed an educational programme specifically designed to prepare them for the extended role (CTN's) and those nurses without such a course (RN's) who were working in similar primary care settings, and (2) recent and non-recent CTN's, in order to estimate the interactive effect of the course and experience following it.

Although the literature is replete with descriptions of both the extended role of the nurse and educational programmes for it, most studies to date have focused on the acceptability of these new professionals. Performance evaluation studies, few in number, have been singularly lacking in documentation of their reliability and validity. A major requirement of any proposed design would be the development of evaluation instruments, sampling procedures, and experimental design controls which would yield results of known reliability and validity. This study was undertaken to produce such a research design for the field evaluation of nurses performing in an extended role in primary care.

Description of the Study

The bank of objectives (cf. p. 2) delineated the skills and knowledge which were to be measured by the evaluation instruments developed. While the content of the criteria was valid, the reliability and validity of the instruments used to measure the content were developed by determining: (1) the consistency of various professionals (physicians, nurses, nurse practitioners, and a CTN nurse-educator) in assessing skills using the instruments, and (2) the statistical significance of differences in scores following the intervention of the CTN course. The sample available for studies of the validity of the instruments was one University of Alberta class of CTN students. All eight of these registered nurses were Medical Services Branch employees with experience in nursing stations prior to course enrollment.

The project also involved the specification of a design for field evaluation of graduates of CTN courses employed in nursing stations by Medical Services Branch. That agency made available morbidity and mortality data, as well as information regarding the accessible population. Additionally, results of pre- and post-tests of all previous classes of CTN students at the University of Alberta could be used in developing a sampling design which would control threats to generalizability of results of a field evaluation. Data from these sources were analyzed, and the results taken into account in the specification of the design for field evaluation.

Assumptions and Limitations of the Study

The findings of the study are specific to CTN graduates, and therefore not generalizable to all nurses in extended roles. Furthermore, there are certain constraints as to the generalizability of the study results for practicing CTN graduates. These considerations include the fact that a small sample (eight) was evaluated at one academic institution rather than in the field. Additionally, it was necessary to assume that:

a) the sample was, prior to the course, representative of control nurses, and upon completion, of the target population of CTN graduates;

b) the scores achieved by the sample CTN's approximate those which would be achieved by recent CTN's in the field (that is, no systematic difference between the testing situation and the field would be observed);

c) the bank of objectives, not yet published but available to all CTN programmes, had been integrated into the curricula of the other universities to the same extent as at the University of Alberta; and

d) the physicians, nurse practitioners, nurses, and CTN nurse-educator who were judges were representative of their professional groups.

Should any of these assumptions be incorrect, results of this study may not be generalizable.

A last concern is with the morbidity data used, in which there appeared to be certain irregularities. This investigator had no basis

for determining the validity of the data, but made the assumption that the data were sufficiently accurate to contribute to the evaluation design.

Definition of Terms

Specific terms and connotations which may be peculiar to this study are listed below:

CTN--a graduate of a Clinical Training for Nurses course (cf. p. 2), as opposed to an RN (registered nurse who has not been enrolled in a CTN course);

non-recent CTN--a nurse who completed the CTN course at least nine months prior to field evaluation, as opposed to a recent CTN (one who completed the CTN course at least three, but less than nine months prior to field evaluation);

paediatric patient--(one who is) less than sixteen years of age, as opposed to an adult patient (at least sixteen years old);

pre-test--a knowledge and/or skill examination taken prior to a CTN course or exposure to the bank of objectives, as opposed to post-test (following both the course and exposure to the bank of objectives).

Format of the Thesis

In the following chapter, literature pertinent to the evaluation of nurse practitioners is reviewed. In Chapters III and IV, the methodology of the study and results of data analysis are presented, both chapters being divided into sections consistent with the major foci of the study: evaluation instruments, and research design. The concluding chapter contains a summary of the study and recommendations arising from the project.

CHAPTER II

REVIEW OF SELECTED LITERATURE

The literature reviewed for this field evaluation design of nurses in an extended role focuses on four areas, namely: evaluative studies of nurse practitioners, reliability of measurement instruments, validity of measurement instruments, and validity in research design.

Evaluation of Nurse Practitioners

Introduction

Although the development of the formalized nurse practitioner role is recent in North America, the health care literature is replete with publications on the subject. The intent here is to provide not an exhaustive review, but a background for the reader unfamiliar with the nurse practitioner, and an overview of evaluation to date.

Two widely recognized definitions will be used in this literature review, which focuses on performance evaluation of nurse practitioners in primary care. A nurse practitioner has been defined by Spitzer and Kergin as:

> a nurse in an expanded role oriented to the provision of primary health care as a member of a team of health professionals, relating to families on a long-term basis and who, through a combination of special education and experience beyond a baccalaureate degree or a diploma, is qualified to fulfill the expectations of this role. (p. 992)

Chioni and Panicucci (1972) have defined primary care as:

> the usual point of entry into the health care system; it is oriented toward the promotion and maintenance of health, prevention of disease, and care of individuals with common

> health problems, uncomplicated illness, chronic latent illness, and selected aspects of complicated illness in the home or out-patient setting. Care is given on a family basis with professionals providing guidance in the use of health resources and referring to other levels of the health care system. (p. 33)

Historical Antecedents

Of the many conceivable arguments, it would appear that the most generally accepted rationale for the development of medical paraprofessionals has been to "increase the productivity of the existing physicians by relieving them of routine ordinary responsibilities and activities which were formerly considered to be the practice of medicine" (Letourneau, 1968, p. 55). In primary care, allocation to paraprofessionals of traditional physician tasks has been a global trend, with considerable variation in personnel used and their training (excellently described in an early article by Connelly, Stoecklu, Lepper & Farrissey, 1966, and by Sidel, 1968).

Among Canadian health professionals it was agreed that the professional nurse is the best person to supplement the physician in primary care (Boudreau, 1972; Mussallem, 1971). As related by Spitzer and Kergin, there has been "no serious challenge" (1973, p. 991) to this approach. This consensus followed a period of considerable difference of opinion in the medical and nursing professions in North America as to the most appropriate type of personnel for this primary care role, and the educational programmes required for them (Day, Egli, & Silver, 1970; Hacker, 1969; Ingles, 1968; Stead, 1966).

These sources are only a small portion of the growing body of literature documenting the introduction of medical paraprofessionals, commonly nurses in extended roles, as providers of primary health care.

Indirect Evaluations

Considerable emphasis in the literature has been placed on descriptive studies of the evaluation of the degree of acceptance by the public of nurses in the extended role. The level of acceptance has been reported as being high in studies of consumers who have had experience with nurse practitioners (Day, Egli & Silver, 1970; Lees, 1973) as well as those who have not (Chenoy, Spitzer & Anderson, 1973).

There are also numerous articles descriptive of doctors' and nurses' attitudes to extension of the nurse's role in primary care (Chioni, 1971; Clelland, 1972; Dolan, 1973; McCormack, 1974; Moore, 1974), but few empirical studies. Reed and Roghmann (1971) did report a study in which attitude questionnaires were presented to medical students before and after their senior year, nurses, and house staff physicians. These professionals were ranked in terms of their degree of acceptance of the expanded role concept. The results showed that the nurses were the most accepting of the concept, followed by the medical students at the beginning of their fourth year, the medical students at the conclusion of their fourth year, and the house staff. Further stratification of these professionals revealed that, "there may be a curvilinear relationship between rank in the hospital hierarchy and acceptance" (p. 375). Although considerable detail was included in that report, significance levels were not stated, nor is it clear whether or not nurses in expanded roles were either part of the sample or of the respondents' experience.

Integral to the continued development of a new role is job satisfaction of those associated with the role (Linn, 1975). In a Canadian study (Spitzer, Kergin, Yoshida, Russell, Hackett & Goldsmith,

1973) of medical practices with nurse practitioners and a control group of practices without nurse practitioners, there was only one area of satisfaction in which significant differences were found. As had been anticipated, physicians in the experimental group were less satisfied with remuneration, as their income was negatively influenced by the nurse practitioners performing some tasks in ambulatory care.

Spitzer et al. dealt with job satisfaction of the nurse practitioners and their associates, Bullough (1974) with that of the practitioners themselves. Bullough put forth the following two hypotheses.

> I. Nurses will find more intrinsic job satisfaction in the extended role than in the traditional nursing role.
> II. These increased intrinsic rewards will be important enough to the nurse practitioners to increase their overall feelings of satisfaction with their jobs. (p. 16)

By questionnaire, 56 nurses (18 of them working in extended roles) and 17 formally trained paediatric nurse practitioners at the beginning and end of their preceptorships were sampled. Work satisfaction and overall job satisfaction measurements were combined with a semantic differential scale. Although "there were significant differences in the levels of intrinsic job satisfaction between the nurse practitioners and the other registered nurses . . . examination of answers to the questions about overall satisfaction revealed no significant differences" (p. 17). That is, the first hypothesis was upheld, but the second was not. The author identified several factors which may have influenced these findings, and stressed the need for further research.

Detailed descriptions of nurse practitioners' legal constraints, responsibilities, and functions have not been forthcoming in Canada, with scant mention in the American literature (Hershey, 1973; Hospital Week, 1973; Krever, 1973). Several authors (Boudreau, 1972; Chioni,

1971; Clelland, 1972; Lambertson, 1971) have supported the position that roles will develop gradually and individually, and might be unduly restricted by such documents. Others (Flynn, 1972; Spitzer & Kergin, 1971) expressed the opinion that minimal behavioral criteria are central to the development of the role. An understanding of the functions which extended role nurses are assuming in primary care can nevertheless be gained from the numerous articles describing their activities (Kergin, Yoshida & Tidey, 1971; Silver & Duncan, 1971; Smale, 1971; Spitzer, Kergin, Yoshida, et al., 1973)

The introduction of paramedical physician extenders has given rise to a great many descriptive articles. As the role has not been universally legitimized, either in terms of public opinion or legislation, much of the literature to date describes acceptance, job satisfaction, functions, and legal constraints associated with the role (be it of nurses with extended roles, or other paraprofessionals). While few in number, some performance evaluation studies have been reported. The following sections discuss literature reporting performance evaluation studies upon course completion and in the field.

Education Programmes' Performance Evaluation

Faculty of the McMaster University programme have outlined their three types of evaluation: written, evaluation of history-taking and physical examination in a clinical setting, and problem solving in video-taped situations. It is noteworthy that pre-set criteria were adhered to for successful completion of the programme (Kergin & Spitzer, 1972).

Written examinations were used by the St. Louis University, which

trains physicians' assistants as well as paediatric nurse practitioners (deCastro & Rolfe, 1974). In contrast to McMaster University, no mention was made in this publication of minimal acceptable scores. Rather, students were compared to physician's assistants, paediatric residents, and senior medical students at the end of their ambulatory paediatric rotation. There was no significant difference among the first three mentioned groups, all of which scored significantly higher than the medical students.

A rigorous evaluation of a programme to train nurses for an extended role in internal medicine was conducted by Flynn (1972).

> The problem under study involved the construction and/or application of evaluative devices to three areas of study: a description of the training program, the effectiveness of the program, and the acceptance of the professional in the extended role. (p. v)

Several techniques were utilized to measure effectiveness of the programme, including one objective and three essay knowledge tests, and measurement of skills in physical examination and history-taking. Flynn, having established the reliability and validity of most of the instruments, concluded that, "the nurse clinicians demonstrate varying levels of ability or accomplishment in fulfilling their original objectives" (p. 233).

It can be anticipated that the literature describing evaluation of students upon completion of educational courses for nurse practitioners will expand with the continued development of formal educational programmes. At present there is a dearth of such literature, with current authors decrying the need for standards and criteria of care (Bailit, Lewis, Hockheiser & Bush, 1975; Williams, 1975). The reader will note also the absence in the literature of criteria for

field evaluations of nurse practitioners, discussed in the following section.

Performance Evaluation in Practice

As was indicated previously (cf. p. 7), evaluative designs and studies of field performance are of central interest. A major activity transferred to the primary care nurse practitioner is patient assessment. Evaluation of nurses' skills in this area is, therefore, of great importance. Comparison of the physical findings of paediatric nurse practitioners with those of a paediatrician was undertaken by Duncan, Smith and Silver (1971). "In an unselected consecutive series of charts, 182 children were first seen by the nurse and subsequently by a paediatrician" (p. 1170). In only 7% was there a significant difference ("defined as applying to an incorrect assessment or a failure to recognize a condition which could be favorably affected by treatment" [p. 1171]) between the assessments of the nurses and paediatricians. Although the study was fairly well documented, readers were not informed as to:

i) whether or not the paediatrician had the nurses' charting available to him at the time of his assessment;

ii) whether or not the paediatrician's opinion was subsequently validated by either another physician, or laboratory findings (except in one case); or

iii) the time span between the examinations of the paediatrician and nurses (there was evidence that the paediatrician's examinations were delayed in some cases for as long as several weeks after the nurses').

The latter two points are important, in that physical findings can change rapidly, particularly in children.

Such a report is still far superior to numerous others with respect to evaluation. For example, a Kaiser-Permanente training programme in physical assessment and history-taking was described in detail in a 1974 publication. The only reference to performance evaluation was ". . . we are not aware of any significant errors of omission in over 30,000 patients who have been through the HAE [Health Appraisal Examination]" (Henrique, Virgadamo & Kahane, 1974, p. 52).

Perhaps of greater importance than the process of delivering care are the outcomes of methods of health care delivery. The literature does describe the following experimental and quasi-experimental attempts to evaluate the outcomes of nurses' practice in the extended role in primary care.

A Scottish triage-like study (Moore, Barber, Robinson & Taylor, 1973) dealt with the decision-making ability of a nurse (not trained in the extended role) regarding the urgency and type of medical care needed by patients seen at home. There was no statistically significant difference between the nurse and three doctors in actions chosen or urgency score assigned. However, the authors stated that while the differences in urgency scores were not statistically significant, they may have been clinically significant, and concluded that "training programmes can be developed for the nurse aimed at eliminating specific risks" (p. 819).

Catalanello, Mingo and Pinches, the authors of "Evaluative Research Design for a Health Manpower Innovation," stated that "the design structure being proposed will allow an evaluation of whether the health

status has, indeed, been effected [sic] by the demonstration-research project under consideration" (1972, p. 229). Although the design involved assessment of such variables as physician extenders' utilization, economic impact, and functions, no evaluation of patients' health status was mentioned. Therefore, the unstated assumption that such variables affect health must be accepted for the design to serve its intended purpose of measurement of output in terms of health status.

In Lewis' (1969) often referred to project, process and outcome were both evaluated, using primarily a critical-incident technique. Chronic clinic patients in a "relatively stable phase of the natural history of their disease" (p. 646) were stratified according to such factors as age, sex and diagnosis before being randomly assigned to care by either a traditional or nurse-operated clinic. At the end of a year,

> although there were no differences in terms of deaths or severity of disease between the two patient groups (nurse clinics and control groups) there were statistically significant differences in outcomes in terms of reduction of disability and relative decreases in discomfort and dissatisfaction of patients seen in the nurse clinic. (p. 648)

A recent publication by a McMaster group (Spitzer, Sackett, Sibley, Roberts, Gent, Kergin, Hackett & Olynich, 1974) described a randomized trial in private practice of 296 patients seen by two physicians, and 521 seen first by two nurse practitioners. Outcome measures were taken of physical, emotional, and social function, although only physical function was evaluated by both pre- and post-testing. Quality of care was also evaluated through two methods: the first "based on identifying and assessing the manner in which the practitioners managed a series of 10 indicator conditions [and the second on] . . . . the manner in which 13 common drugs were prescribed" (p. 253). Some limitations of the study were recognized by the authors,

whose research design was detailed in one publication (Spitzer et al, 1974), while results were disclosed in another (Sackett, Spitzer, Gent & Roberts, 1974). The key issue in the analysis is that, although Spitzer et al described "the role of 'experimental subject,' [as being] assumed by the collaborating physicians and nurses" (1974, p. 255), the patients were actually the subjects of the experiment. Therefore, while predictions to the patients of these two nurse practitioners can be made, it is not legitimate to generalize the results of this study to the care provided by all nurse practitioners.

Summary

The nurse practitioner role is a relatively recent development in health care delivery. As Williams (1975) stated:

> most studies of the nurse practitioner have focused primarily on developing information used in making decisions related to the employment of nurses in expanded roles and to the development and evaluation of training programmes. (p. 176)

The literature suggests that expanded role nurses in primary care are: acceptable to the public and health professionals, satisfied with their new role, and performing diversified functions in a wide variety of ambulatory settings.

Few educational programmes for nurse practitioners have published reports of methods used in evaluation of graduates' performance. Evaluation of practice has also been limited, generally comparing performance or outcome with that of physicians, rather than to norms or pre-set criteria for nurse practitioners. Much analysis to date has been of opinion, not nurse practitioners per se, and there are infrequent reports of the reliability and/or validity of the measurement instruments, the subject of the literature review which follows.

Reliability of Measurement Instruments

Reliability may be defined (American Psychological Association, 1966) as the "accuracy (consistency and stability) of measurement by a test" (p. 25). The reliability of a measure, as indicated by either the standard error of measurement or the reliability coefficient (Thorndike & Hagen, 1961), represents that proportion of test variance free from measurement error variance (Kerlinger, 1973; Cronbach, 1960). In general, reliability increases with the amount of behavior sampled, although Cronbach pointed out that extremely long tests can lead to boredom, and therefore a reduction in reliability.

Reliability is dependent upon agreement between two maximally similar methods of measuring the same trait (Campbell & Fiske, 1967). Methods of obtaining repeated measures estimates (Ebel, 1951; Cronbach, 1960; Thorndike & Hagen, 1961) include:

a) repetition of the same test,

b) administration of a second equivalent form of the test,

c) subdivision of the test into two or more equivalent fractions,

d) averaging all possible split-halves (alpha coefficient for continuous data and the Kuder-Richardson formulae for dichotomous data), and

e) rating performance by two or more judges.

Each method of assessing reliability is subject to some limitations, related to either the identification of sources of variance, or to more pragmatic problems. Numerical methods of establishing reliability estimates also differ.

Ebel (1951) discussed at length the development of inter-rater

reliability estimates. This method is appropriate when two or more measures other than test scores are available, these often being performance ratings. In such instances, the investigator must be selective in employing formulae for estimating the reliability, using the correlation coefficient only when either the means and variances of the judges are considered equal, or the scores are standardized to the same mean and variance. Maguire and Hazlett (1969) extended upon Ebel's work and that of Cronbach, Rajaratnam and Gleser (1963) in their demonstration of reliability estimation in instances when differences in (1) means and variances, (2) means only, and (3) neither means nor variances, are of interest. In the last situation, the authors point out, the correlation coefficient is appropriate following standardization of scores to the same mean and variance, whether the reliability being estimated is inter-judge or another measure of internal consistency.

Both the formula for, and method of establishing reliability are dependent upon the nature of the test and/or the research design, and must be selected accordingly. While reliability is a necessary quality in a measurement instrument, the sufficient condition is its validity. The various types of validity, as described in the literature, are discussed in the next section.

Validity of Measurement Instruments

Introduction

> The validity of a measuring instrument may be defined as the extent to which differences in scores on it reflect true differences among individuals, groups, or situations in the characteristic which it seeks to measure, or true differences in the same individual, group, or situation from one occasion to another, rather than constant or random errors. (Selltiz, Jahoda, Deutsch & Cook, 1959, p. 155)

The definition of this important concept in measurement has been more concisely put by Thorndike and Hagen (1961) as "the extent to which a test measures what we actually wish to measure" (p. 160). In the education literature, the need for validity in measurement instruments has been expressed by many authors.

While reliability is a necessary quality in a test, it cannot supplant validity, as a test may dependably measure behavior other than that intended. Conversely to reliability, validity rests upon the convergence of independent (rather than similar) methods of assessment.

Tests or items therein cannot be considered valid or invalid, nor can their absolute validity be declared. Selltiz et al (1959) clearly made the point that, since a person's true measure on a variable is never known, absolute validity cannot be determined. "The validity of a test is clearly a matter of degree, not an all or none quality . . . . [Tests] are more or less valid" (Ebel, 1972, p. 447). Such restriction on the quantification of total validity is due both to the absence of perfect criteria and to the inherent multiplicity of types of validity. Although some kinds of validity lend themselves to numerical estimation, others are not usually considered quantifiable. Apropos of this, Ebel classified types of validity as being either direct or indirect, while Thorndike and Hagen (1961) referred to them as those dependent, respectively, on professional judgement and on rational analysis. The following discussion deals first with two types of validity commonly referred to as judgmental, face and content, and then with those for which empirical estimates are usually obtained: concurrent, construct, and predictive.

Face Validity

Face validity has been variously defined as the appearance of reasonableness of a test (Thorndike and Hagen, 1961) or "what the test appears to measure" (Ebel, 1972; Fox, 1970). Thorndike and Hagen's reference (in a footnote only) stated that, "what a test 'looks like' may be of importance in determining its acceptability and reasonableness to those who will be tested" (p. 164). While these authors referred to face validity as being judged by tested subjects, they did not discuss face validity in terms of those administering the test.

Mosier (1967) strongly argued that the term face validity should be abandoned. He took the position that there are the following three meanings attributed to the term:

> a) [Validity by assumption] The test bears a common-sense relationship to the measurement objective and therefore no statistical verification is necessary . . . .
> b) [Validity by definition] The test sets such a task that the universe of possible tasks (of which the test is a representative sample) is the only practicable criterion and the test is therefore a valid measure of the universe defined in terms of the sample. This implies that the test is a valid measure of whatever trait is measured reliably by the test . . . .
> c) [Validity by appearance] In the interest of the acceptability of the test to those most intimately concerned with its use, it is highly desirable that a test possess not only statistical validity, but also, as an added attribute, the appearance of practicality. (p. 218)

Mosier propounded that the assumption of validity is fallacious or even dangerous, and should never be made. Selltiz et al. (1965) supported this position. A refinement of Mosier's "validity by definition" might be regarded in contemporary literature as closely related to content validity.

Validity by appearance was considered important by Mosier. He felt that the test should be acceptable to the consumers, including

both persons taking the test, and persons using it. "The appearance of practicality is an objective sufficiently desirable in its own right that it may often be sought as an additional end consistent with the principal objective--predictive value" (p. 215). While omitted in other authors' discussions of validity (Anastasi, 1968; Cronbach, 1960), this "validity by appearance," or face validity, may perhaps be of considerable contemporaneous importance with the trend to contracted research, as "if a test is to be used effectively in achieving its objectives, it is essential that it actually be selected for use" (Mosier, 1967, p. 214).

Content Validity

Content validity is "established by showing that the test items are a sample of a universe in which the investigator is interested" (Cronbach & Meehl, 1967, p. 245). Content validation rests essentially on authorities' judgement of the representativeness of the sampling of the universe, the universe itself being as clearly defined as possible. Even though the content of the universe to be tested may have been defined, Kerlinger (1973) cautioned that "the items of a test must be studied, each item being weighed for its presumed representativeness of the universe . . . weighed for its presumed relevance to the property being measured" (p. 459). Ebel (1966) further insisted that the directions for administering and scoring the test should be subjected to logical analysis and expert judgement, as should the rationale and specifications for the instrument. Only after all these judgemental assurances of validity have been received, can content validity be assumed.

Concurrent and Predictive Validity

Two measures of validity which are estimated empirically are concurrent and predictive validity. Both of these are estimated by the degree of consistency which a test has with a criterion measure, usually expressed by a correlation coefficient. Selltiz et al. (1959) classified both as pragmatic in that some other behaviour or characteristic is being predicted; the term 'empirical' was used by Thorndike and Hagen (1961) in reference to the statistical estimation of these validity estimates. More commonly, both are described as being criterion-oriented, as the principal concern is with the relationship between the test scores and scores on an independent criterion (Cronbach & Meehl, 1967; Kerlinger, 1973).

Differentiation between concurrent and predictive validity is primarily on the basis of time, concurrent being established by comparison of the instrument with the criterion at the same time, predictive validity by comparison with a criterion measure in the future (Cronbach and Meehl, 1967). Much of the literature gives the impression that more than time dimension separates these two types of validity, but does not clarify the issue. Various examples indicate that concurrent validation is often based on a criterion measuring essentially similar behaviour, predictive validation based on different behavior, in that performance on a related issue such as job or academic success is predicted from an earlier measure of aptitude, intelligence, achievement in other academic work, etc. (Cronbach, 1971; Sjoberg & Nett, 1968). Cronbach (1960) wrote lucidly on this subject, however, when he described predictive validity as a record of outcome, and stated:

> Concurrent validity is investigated when the test is proposed as a substitute for some other information; this information is then the criterion. Designers of new tests frequently establish concurrent validity for their instruments by comparing them to established tests. (p. 109)

"The single greatest difficulty of criterion-related validity is the criterion" (Kerlinger, 1973, p. 460). Thorndike and Hagen (1961) discussed four qualities desirable in a criterion measure, the first being that the measure be relevant. The degree of relevance being unmeasurable, the authors stated it is "necessary to rely upon professional judgement to provide the appraisal of the degree to which any partial criterion measure is relevant to the ultimate criterion" (p. 166). A criterion measure should also be free from bias, e.g., provide similar scoring opportunities for all persons. The third quality the measure should have is reliability, defined as being "stable or reproducible" (p. 167). Finally, the criterion measure should meet the practical considerations of convenience and availability.

Construct Validity

"Scientifically speaking, construct validity is one of the most significant advances of modern measurement theory and practice" (Kerlinger, 1973, p. 461). Cronbach and Meehl (1967) described the basis of construct validity: rather than on the measuring instrument, the focus is on the measured construct defined as "some postulated attribute . . . assumed to be reflected in test performance" (p. 247). Of concern primarily in the field of psychology, this type of validation is used when no criterion measure of a construct exists, and leads, through inference based on correlations, to an induction. Construct validation can be claimed only when "the interlocking system of laws which constitute a theory [the

nomological network], . . . makes contact with observations, and exhibits explicit, public steps of inference" (pp. 255-256).

Cronbach and Meehl discussed five methods of construct validation, including:

1) testing the expectation, based on the investigator's understanding of a construct, that two groups will differ;
2) obtaining a correlation matrix of two or more tests presumed to measure the same construct, and using factor analysis to divide the construct into a more pragmatic frame of reference, or to lead to a numerical estimate of the bounds of construct validity;
3) studying the internal structure of the test for either positive or negative expected item-test correlations (measures of internal consistency necessary for validity but not indicators of it);
4) examining the stability of test scores over time, with or without experimental intervention (often transient influences); and
5) observing the tested subject's process during his test performance.

Campbell and Fiske (1967) are the principal authors who operationalized the demonstration of construct validity by the establishment of convergent and discriminant validity by the multitrait-multimethod matrix. In this method of validation, the measurement of at least two traits (constructs) by each of at least two methods leads to both reliability and validity estimates. Convergent validity may be said to be shown when two different methods of measuring the same trait correlate significantly. The authors discussed three criteria which provide evidence of discriminant

validity, that is, that no preponderance of method variance is present. Numerous matrices were interpreted to illustrate this approach to construct validity, the limitations of which were also indicated by Campbell and Fiske.

Regardless of the method of construct validation used, validation supports both the test and the network of theory. If validity is not demonstrated, both the theory and instrument must be questioned.

Summary

The types of validity of instruments decribed frequently in the literature are: face, content, concurrent, predictive, and construct. Face and content depend on judgement of the acceptability of an instrument in terms of, respectively, the apparent reasonableness of the test, and the degree to which the instrument samples the universe under consideration. Correlation of test scores with those on a criterion administered at the same time leads to concurrent validity, correlation with a criterion in the future to predictive, both yielding numerical estimates. Construct validity is established when the measurement of a theorized construct is demonstrated, preferably by establishing evidence of convergent and discriminant validity with the multitrait-multimethod matrix.

Validity, specific to the situation under which validation studies were conducted, is not applicable to all possible uses of an instrument. Inasmuch as no limits have been set on acceptability of levels of validity, and some types of validity do not lend themselves to numerical estimates, the test developer should establish the maximal possible validity for an instrument, and report both the procedures and results of validation studies.

Design Validity

Introduction

Four general research methodologies are found in the literature--historical, descriptive, quasi-experimental, and experimental. Designs for contemporary studies in the behavioural sciences are often of the last two types. Campbell and Stanley (1963) described an experimental design as being one "in which an experimenter having complete mastery can schedule treatments and measurements for optimal statistical efficiency, with complexity of design emerging only from that goal of efficiency" (p. 1). In quasi-experimental designs, the experimenter lacks this complete control.

A basic understanding of the strengths and weaknesses of various research designs can be gained from the publications of Campbell and Stanley (1963), and Bracht and Glass (1968), major contributors to the literature on internal and external validity in research designs. Regardless of the methodology used, maximal internal and external validity in design should be achieved.

Internal Validity

Campbell and Stanley stated that internal validity answers the question: "Did in fact the experimental treatments make a difference in this specific experimental instance?" (p. 5). The following, drawing heavily on their work, is a list and brief discussion of extraneous variables which may reduce the internal validity of research findings:

a) history, meaning specific change-producing events which have taken place in addition to the experimental variable;

b) respondent maturation, or change biologically and psychologically between testings;

c) pre-testing experience leading to changes, usually improvement, in post-testing scores;

d) instrumentation, or changes in the measuring instrument, observers, or raters;

e) statistical regression, that tendency of groups selected for their extremity to gravitate toward the population mean;

f) selection of comparison groups differentially rather than randomly;

g) experimental mortality leading to differential reduction in the groups being studied; and

h) interaction of the previously listed factors.

Unless there is reasonable control of these factors, the experimental variable cannot be assumed to be reflected to an adequate extent in research findings. While both internal and external validity are desirable, the latter is not possible without the former, as is seen from the following section.

External Validity

External validity was defined (Cook, 1969) as being present when generalizations can be made from the experimental situation to non-experimental situation which the former is said to represent" (p. 210). In generalizing to the non-experimental situation, consideration should be given to treatment and measurement variables, populations, and settings.

Factors jeopardizing external validity (as listed by Campbell and Stanley) include the reactive or interaction effect of pre-testing,

interaction effects of selection biases and the experimental variable, reactive effects of experimental arrangements, and multiple-treatment interference. In the Bracht and Glass (1968) explicatory article, these threats, in addition to others not listed by Campbell and Stanley, were divided into two broad classes, "correspond[ing] to two types of external validity: population validity and ecological validity" (p. 439).

Two major threats to population validity exist. The first of these is the difficulty generalizing from empirical studies to the total population under study. As Bracht and Glass (1968) stated,

> The experimenter must make two 'jumps' in his generalizations: (1) from the sample to the experimentally accessible population, and (2) from the accessible population to the target population. The first jump, a matter of inferential statistics, usually presents no problem if the experimenter has selected his sample randomly from the accessible population. (p. 440)
>
> The second jump, from the experimentally accessible population to the target population, can be made with relatively less confidence and rigor than the first jump. The only basis for this inference is a thorough knowledge of the characteristics of both populations and how these characteristics interact with the experimental treatment. (p. 441)

The second factor which may reduce population validity is the interaction between some personological variable of the subjects and the treatment variable. If this interaction is statistically significant, further investigation is required to discover whether generalizability is limited, or if "one treatment can be prescribed for all levels of the personological variable" (p. 444).

Threats to ecological validity include the following:

a) lack of sufficiently explicit description of the independent variable to permit replication of the study and/or estimation of the generalizability of results to other situations;

b) consecutive administration of treatments preventing estimation of the effect of other than the first treatment;

c) subjects' response to the knowledge that they are participating in an experiment (known as the Hawthorne effect, and discussed at length by Cook in 1969);

d) novelty and disruption effects of new treatments, leading respectively to temporarily improved and reduced effectiveness of the treatment;

e) reactions of the subjects to the experimenter, known as observer effect or experimenter effect;

f) sensitization of subjects to the treatment by pre-testing;

g) sensitization of subjects to the treatment by post-testing;

h) the effect of administering the treatment under particular historical conditions;

i) lack of sufficient specificity in the operation definition of the dependent variable, to allow identification of the variable under study when treatments may have a multiplicity of outcomes; and

j) variability in residual effect of the treatment over time.

Summary

A design is said to possess internal validity when the findings reflect accurately the influence of the independent variable. The internal validity of various research designs may be threatened by extraneous variables, as explicated by Campbell and Stanley (1963).

If the results of an experiment can be relied upon, generalization from them may be possible. There are also numerous threats, as have

been described by Bracht and Glass (1968), to the external validity or generalizability of research designs.

Although the reduction of sources of internal invalidity may increase the threats to generalizability (and vice versa), it is important that the maximal degree of both internal and external design validity be achieved, whether the research is experimental or quasi-experimental in nature.

Summary of Chapter

In this chapter, selected literature has been reviewed on the subjects of nurse practitioner evaluation, and validity in measurement instruments and research design.

The few performance evaluation studies of nurse practitioners documented in the literature have frequently been comparisons with physicians rather than with criteria for their own professional group. Reports of validation procedures for the measurement instruments used in evaluating nurse practitioners are negligible, although instruments of known reliability and validity are considered obligatory by authors in the field of measurement. The various types of validity have been reviewed (face, content, concurrent, predictive and construct), as have internal and external validity in research design. In the evaluation of primary care paramedical physician extenders, some attention to validity have either been inadequately controlled, or their control inadequately described.

In the following chapter, the validation studies of the instruments central to this project are presented, as are the methods used to ensure

representativeness of the target population if a sample of nurse practitioners (CTN's) and their control group are evaluated in the field.

CHAPTER III

METHODOLOGY

Introduction

Two major areas of investigation were undertaken in the project. The primary focus was the development of instruments of known reliability and validity for evaluation of representative skills and knowledge expected of CTN's. Secondary was the design for on-site evaluation incorporating the measurement instruments. In the research design, it was important to achieve reduction of the effects of sources of internal and external invalidity while maintaining a cost-effective format. The instruments and sampling design, together with an instruction manual and recommendations for implementation comprised a package for field evaluation of CTN's.

Figure 1 illustrates the overall project design of this study. The methodological description provided in this chapter corresponds to each of the major areas of investigation noted in Figure 1.

Validity in Measurements

Criteria in Selection of Objectives

The content-valid bank of behavioural objectives (cf. p. 2) was used as the basis for the evaluation package, thus assuring representativeness of the skills and knowledge expected. There was no a priori stratification in terms of degree of importance of these objectives--all were considered

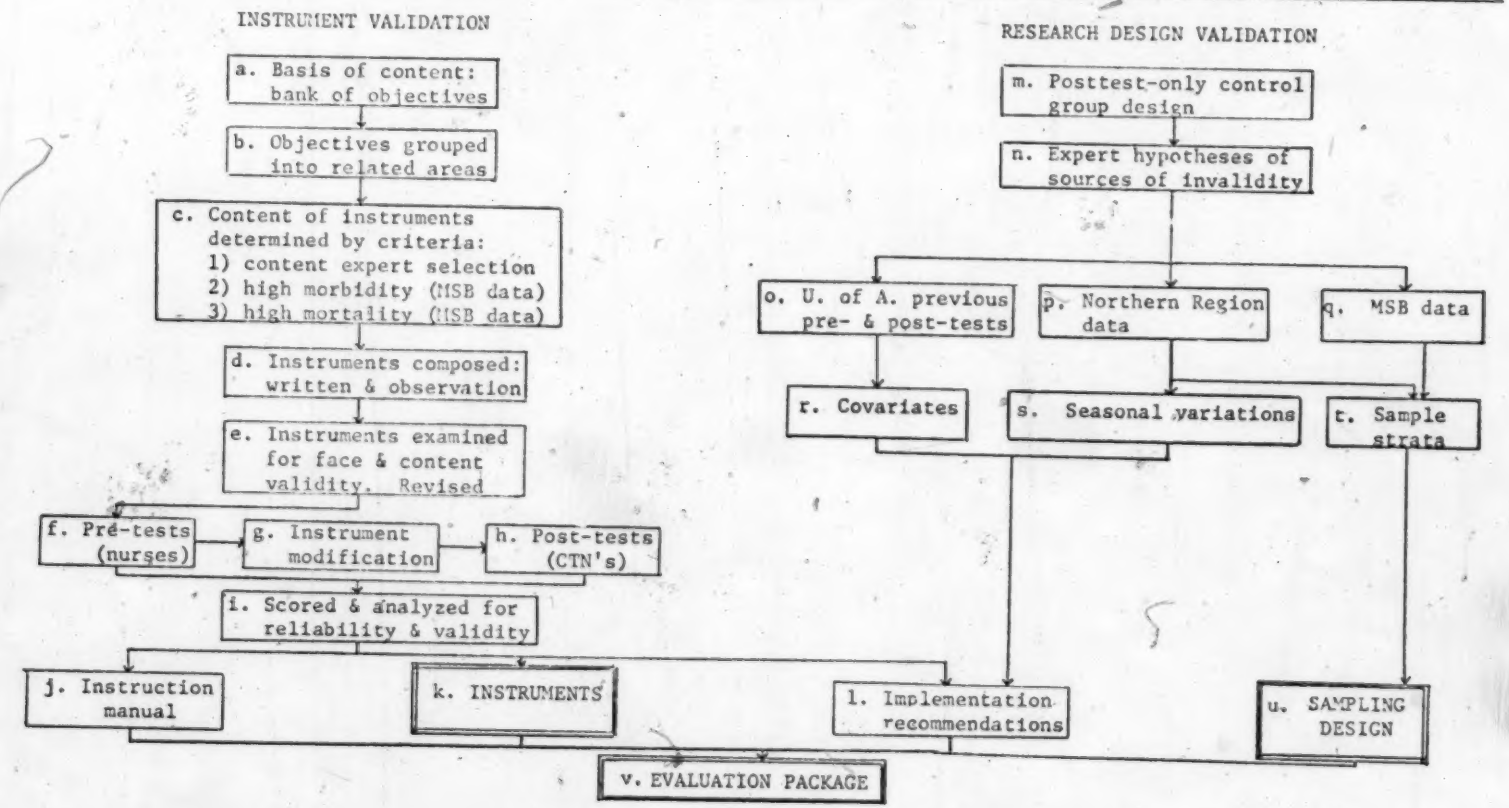

INSTRUMENT VALIDATION
a. Basis of content: bank of objectives
b. Objectives grouped into related areas
c. Content of instruments determined by criteria:
1) content expert selection
2) high morbidity (MSB data)
3) high mortality (MSB data)
d. Instruments composed: written & observation
e. Instruments examined for face & content validity. Revised
f. Pre-tests (nurses)
g. Instrument modification
h. Post-tests (CTN's)
i. Scored & analyzed for reliability & validity
j. Instruction manual
k. INSTRUMENTS
RESEARCH DESIGN VALIDATION
m. Posttest-only control group design
n. Expert hypotheses of sources of invalidity
o. U. of A. previous pre- & post-tests
p. Northern Region data
q. MSB data
r. Covariates
s. Seasonal variations
t. Sample strata
l. Implementation recommendations
u. SAMPLING DESIGN
v. EVALUATION PACKAGE

Figure 1. Illustration of methodology of development of evaluation package. (MSB refers to Medical Services Branch.)

"necessary" (Hazlett, 1975). Random selection of individual objectives from this universe would have been possible as their specificity does allow each to stand alone. Such randomization was rejected, however, in favor of the development of instruments which measured areas of related skills and/or knowledge. Thus the instruments were systematically organized, reflecting the continuity inherent in the processes of patient assessment and care. Random selection of objectives grouped into related areas was rejected as it was hypothesized that: (1) content experts would consider some areas of the objectives more important for evaluation, and (2) some behaviors in the objectives would be so infrequently required in nursing stations that they would be unlikely to be observed during field evaluation. An area was therefore selected for evaluation if it met at least one of the following criteria:

1) judged by content experts as being of highest priority for evaluation;
2) related to disease categories of high morbidity; and
3) related to disease categories of high mortality.

Accordingly, the objectives were grouped (cf. Figure 1.b) into areas of common content and presented to workshop participants from CTN universities and Medical Services Branch (cf. Figure 1.c.[1]) with the request that evaluation of the areas be rated as being: necessary, desirable but not necessary, or unnecessary. Eight validators participated, four of whom were CTN physician-educators, two were CTN nurse-educators, and two were nursing officers with Medical Services Branch. Those areas which were used in evaluation were those which at least 75% of the validators indicated must be evaluated.

To utilize the second criterion (that objectives selected be related to conditions causing high morbidity), an analysis was made (cf. Figure 1.c.[2]) of all reported patient visits in 1973 to nursing stations in Medical Services Branch's Northern Region (N.W.T.). Data for each patient visit included the date and place of visit, the patient's date of birth and coded diagnosis (Northern Region uses 188 diagnostic categories). In order to determine the types of patients seen in nursing stations outside of the Northwest Territories, daily record books were requested from each of Medical Services Branch's regions in Southern Canada which operate nursing stations. Although nine books were received, no comparable diagnostic code was in general use, nor was it possible to match the nurses' entries to the N.W.T. code, due to the lack of specificity in recorded diagnoses. Consequently, the categorized diseases with which at least 1,000 patients were seen in Northern Region in 1973 were those considered to meet the second criterion (representing at least one patient every two weeks per nursing station, barring seasonal or other variations).

As indicated by Figure 1.c.(3), the third criterion for selection of objectives for evaluation was that they pertain to diseases causing high mortality, those categories causing a total of 75% of the deaths in the N.W.T. in the years 1972 and 1973 being considered the leading causes of death for the purposes of this study.

The three criteria described were adhered to in the selection of objectives on which test items were based. The particular areas of objectives meeting with the criteria, indicated in the following chapter, fell broadly into patient assessment and suturing behaviors, and medical, paediatric and obstetrical knowledge. A description of the instruments

composed to measure the performance of selected objectives and the methodology used in their validation follows.

Instruments Developed

Within the areas selected for the evaluation, the objectives require a CTN to: (1) have immediate recall of knowledge, (2) be able to recognize either physical findings or facts, or (3) perform a skill. To test performance of these objectives, the types of instruments which were formulated included: written, observation, record review, and simulation. The written and observation instruments have undergone validation studies and are central to this study (cf. Figure 1.d). To enable replication of results, the evaluation package also included an instruction manual.

Written examinations tested achievement of those objectives requiring a knowledge base with emphasis on the diagnosis and management of disease categories which met selection criteria. The format of the examinations included short answer, true or false, and multiple choice items. To increase reliability in the marking, essays were avoided (Ebel, 1972). While all questions were based on the bank of objectives, the bank's correct responses were supplemented with additional acceptable answers documented in the literature, the bank not being all-inclusive of medical knowledge. The proportion of test items referable to paediatrics approximated the proportion of patients seen in the N.W.T. in 1973 who were less than sixteen years of age. The subject matter for the written examinations was divided into two instruments: Obstetrics and the Newborn, and General and Paediatrics (cf. Appendices A and B).

Observation was considered the most effective method of assessing whether or not skill performance met the conditions of acceptability specified

in the objectives (Ebel, 1972). Therefore, forming a large portion of the evaluation, the several observation instruments were designed to evaluate the degree of mastery of objectives requiring performance of skills. For each item on these instruments, the evaluator was required to make a dichotomous decision as to whether or not the performance met the criterion, thus precluding partial marks being awarded. Both this method of marking and the high degree of specificity inherent within each item were utilized to increase reliability.

As patient assessment received high priority ranking by the expert validators (confirming Kergin's [1970] findings), physical examination and adult and paediatric history-taking evaluation instruments were composed. In the evaluation of patient assessment, standardization of tasks across subjects, essential in valid evaluations, was problematical due to the inherent variation in patients' conditions. The physical examination and history instruments therefore required complete routine patient assessment, rather than that partial examination appropriate to a presenting complaint. Even with the degree of standardization achieved by this format, patient dissimilarities were anticipated to present variability in the on-site assessments. Therefore, the instruction manual described appropriate types of patients for assessment, and included standardizing questions (cf. Appendix G) in an attempt to give equal scoring opportunities to all subjects.

In addition to the history and physical observation instruments, one to evaluate suturing was included, as lacerations caused sufficiently high morbidity that suturing was expected to present as a task during the evaluator's visit. As in the other observation instruments (all to be found in

Appendices C through F), criteria for acceptable performance were clearly specified.

The studies which were undertaken to estimate the degree of validity and reliability of the written and observation instruments are described in the following section.

Validity and Reliability of Instruments

Validation studies of the measurement instruments (cf. Figure 1.e to Figure 1.i) included face, content, concurrent, and construct validity. All instruments were administered as pre- and post- tests to a class of eight CTN students at the beginning and conclusion of their course at the University of Alberta. Hence the subjects formed their own control group (i.e., were considered to be RN's on pre-testing and CTN's on post-testing), and between-person variability was excluded from residual error in data analaysis, yielding a more accurate estimate of error variance (Kerlinger, 1973).

Written examinations. Prior to utilization in the pre-test, the written General and Paediatrics, and Obstetrics and Newborn (latter section) examinations were submitted for content validation to the university's CTN nurse- and physician-educators and a paediatric nurse practitioner intimately familiar with the course objectives. The entire Obstetrics and Newborn examination was critically examined by the CTN nurse-educator and two University of Alberta School of Nursing faculty members with expertise in midwifery (one of whom was teaching obstetrical care on the CTN course). As indicated in Figure 1.e, changes suggested by these validators were incorporated into the instruments.

Pre-testing demonstrated the need for some modification of the test items. The majority of changes which were made were changes in the stems to include medically correct responses which had not been cited in the bank of necessary behavioral objectives. (Texts used in documenting these modifications were Chatton, 1972; Conn & Conn, 1971; Davidson & McLeod, 1972; Hellman & Pritchard, 1971; Houston, Joiner & Trounce, 1968; Hughes, 1971; Kempe, Silver & O'Brien, 1972; Lyght, 1966; Myles, 1971; Oxorn & Foote, 1968; Taber, 1956; and Talso & Remenchik, 1956.) This measure was taken to increase, through improved universe sampling (cf. p. 21), the content validity of the instruments.

Following both pre- and post-testing, the written examinations were marked only by this investigator (a possible source of unreliability), but the scoring was later scrutinized by CTN faculty. Test scores were analyzed for reliability using the alpha coefficient, for concurrent validity by correlating the two tests, and statistically significant changes in means from pre- to post-testing (considered indicative of construct validity [cf. p. 24]).

Observation instruments. In the validation of instruments for evaluation of skills by observation, a major difficulty was the error variance which results from individual patient differences. To the degree to which individuals are unique, each patient confronts the primary care provider with a degree of distinctness in patient assessment. In evaluation, such individual differences can increase error, therefore, by acting as variations in the testing situation. During this study, in order to reduce error variance by presenting similar tasks to all subjects, patient simulations were used for the adult histories and physical examinations, and parent simulations for the paediatric histories. The utilization

of simulated patients thus allowed assessment problems to be held constant across students (although patients differed from pre- to post-testing), and dog lacerations on which suturing was evaluated were made as similar as possible by Medical-Surgical Research Institute personnel.

As consistencies between markers were used in estimating reliability and validity (cf. pp. 40-41), audiovisuals were employed to: (1) obviate having several observers present during the subjects' performances, thus reducing observer effect; and (2) promote independence in marking. Recordings were made of the physical examinations (video-taping employing three cameras), the adult histories (by split-screen video-taping), and the paediatric histories (audio recordings).

The content of the physical examination instrument was validated initially by six nurses and four physicians who were representatives at a workshop of participating CTN universities in September, 1974. Judges who assessed performance on the pre- and post-tests also critiqued the history, physical examination, and suturing instruments which they used in terms of content and face validity. Their comments resulted in some changes in the forms, which also required minor modifications for use in the practice setting.

The eight CTN's who formed the project sample also gave evidence of face validity in an unstructured and unrecorded interview. Their general response to the examinations was that, although they had been moderately stressed, they had been fairly tested on important areas of the curriculum.

The correlation within a professional group of judges (nurses, doctors, etc.) was used as a measure of reliability (cf. pp. 17-18). The analysis of variance model as well as the alpha coefficient were used to obtain reliability estimates. Inter-professional correlations, in view of differing educational

and experiential backgrounds of various health care professionals, were considered dissimilar methods, and therefore treated as estimates of validity for each observation instrument. Furthermore, the professional qualifications required in the evaluator(s) were of pragmatic interest, as they would affect personnel costs if the design were implemented. Accordingly, two physicians, two nurses, two nurse practitioners, and this investigator (a CTN nurse-educator) independently marked all physical examinations and histories from the audio-visual recordings. Additionally, the subjects were rated by a physician during the performance of the physicals. This physician several weeks later judged performance from the video-tapes without knowledge of the total scores he had previously assigned, thus providing some basis for determining the consistency of scores from a recorded performance with those allotted on a live performance.

Proficiency in suturing during validation studies was rated by only one member of each professional group, since the suturing form was a modification of a test previously used by the University of Alberta's CTN programme. The inter-judge reliability estimates of that previous surgery instrument had been calculated on three occasions, and found to be .92, .89, and .95.

At least one of each pair of physicians was considered a content expert as they had specialized preparation in the area in which they were evaluating (e.g., a surgeon rated suturing). However, with the exception of two paediatric nurse practitioners (including this author), none of the nursing personnel had related preparation beyond the CTN course. Although the judges for each pre-test marked the corresponding post-test, not all observation tests were marked by the same raters. A total of

twenty individual judges was used including seven physicians, five nurse practitioners, seven nurses, and one CTN nurse-educator. Judges who took part in the project were not randomly selected, but chosen for their ability and willingness to participate. This lack of randomization reduced generalizability of the validity estimates. However, some evidence of the judges' representativeness could be obtained by examination of the similarities of the inter-professional consistencies on different tests by varying representatives of each profession.

In addition to the inter-professional consistency method for estimating validity, the mean differences on pre- and post-tests were calculated for all observation instruments as estimates of construct validity, as had been done with the written examinations. For the field evaluation, it was also necessary to establish whether the instruments' construct validity was sufficient to discriminate between uncorrelated groups of RN's and CTN's. Therefore, with the subjects' scores treated as results of independent groups, analysis was done for mean differences between the testing occasions.

The investigation took the several forms described, as in the absence of criterion measures for CTN's, validity would best be established with a variety of approaches indicating the quality of the instruments. As well as being theoretically necessary, validation was considered imperative as:

1) the scores achieved by the sample actually formed a large proportion of the final assessment of their achievement on the CTN course; and
2) the instruments were to be incorporated into the design for field evaluation which may be implemented by Medical Services Branch.

A description of the methodology used to maximize the validity of the design for the field evaluation is presented in the succeeding section.

The Design Selected

The literature describes many threats to both the internal and external validity of various research designs (cf. pp. 26-29). Considerations of both types of validity were taken into account in formulating the recommendations for the project design (including sampling), and its implementation in on-site evaluation.

If an on-site evaluation were to be done, there would be two severe constraints on the ease of data collection:

1) the widespread geographical distribution of the nursing stations; and
2) the limited number of nurses posted to any one station (a maximum of three, including CTN's).

In the selection of the design, these factors received careful consideration to ensure practicality while yielding findings which would allow prediction.

The basic design recommended involves evaluation of randomly selected CTN's with a randomly selected control group of nursing station nurses. This design, called the "posttest-only control group design" by Campbell and Stanley (1963, pp. 25-26), is a legitimate experimental design.

Internal Validity

While the "posttest-only control group" design has strong internal validity in the traditional classroom setting, the field assessment of CTN's would pose some threats to internal validity. One limitation would be the inability, due to the limited size of the sample and accessible populations, to control by stratified sampling the effects of subjects' antecedent variates if several were found to influence the

dependent variable. The variates hypothesized to have influence were: (1) type of nursing education, (2) length of time since graduation, (3) midwifery training, (4) length of nursing station experience, and (5) age.

Prior to this project, extensive pre- and post-testing had been carried out at the University of Alberta with instruments which, although not based exclusively on the bank of objectives, were considered by that university's CTN educators to have face and content validity in terms of the curriculum. The data from this sample of thirty-two students were analyzed (cf. Figure 1.n to 1.p) to estimate the influence of the preceding antecedent variates. After standardizing the data to the same mean and variance across judges and classes, the analysis of variance model was applied. The strata selected for the antecedent variates are indicated in Table 1. Documentation being unavailable of appropriate levels, choice of those used was based on their ease in implementation in a sampling design, as well as on discussions with nursing educators and administrators.

Table 1.

Stratification of Antecedent Variates Hypothesized to Influence Performance of RN's and/or CTN's

Variates	Criteria	Strata		
		Level 1	Level 2	Level 3
Midwifery	Course length (months)	0	6	12
Nursing station experience	months	<12	≥ 12	
Age	years	<26	26-30	≥ 30
Nursing education	Course completed	RN	Additional diploma	B.Sc.

Note. Influence of length of time since graduation estimated only by regression analysis.

In that the strata for the antecedent variates were arbitrarily chosen, the step-wise regression model was also applied to both pre- and post-test scores to ensure that the strata used had not given misleading results (Kerlinger, 1973). If the stratification of the continuous data for some of the variates had concealed significant predictors, covariates could be used in analysis of data from an on-site evaluation. Such covariate analysis would also overcome the difficulties imposed on stratified sampling for the antecedent variates by the necessarily small sample size.

External Validity

The external validity of the field evaluation project is threatened by several factors described by Bracht and Glass (cf. pp. 29-30). Some threats to ecological validity would be controlled by the posttest-only control group design. Determination of the interactive effects of time and the experimental variable, a request of Medical Services Branch, would be achieved by the recent/non-recent proportional stratification of CTN's prior to random sampling. As well as determining the effects of time, this sampling technique would eliminate the influence of repeated post-testings, and allow completion of the field evaluation within a limited time period. Calculation of proportions in the accessible populations was based on projections by Medical Services Branch (1974) and the length of employment and attrition rate of fifty-three CTN's sponsored by Northern Region (cf. Figure 1.p, 1.q).

Another extraneous ecological variable hypothesized to influence evaluation results was the size of the nursing station caseload, thought to correlate positively with diagnostic and management skills, and nega-

tively with detailed record keeping. Data for Northern Region were therefore analyzed for the correlation between numbers of patients seen and the community size served by the nursing station, although the reporting of patients seen (cf. p. 35) appeared to be questionable in some instances, an opinion supported by the region's senior nursing officer.

While the measures which have been described were taken to ensure an acceptable level of validity in measurement and design, other possible sources of invalidity are discussed in Chapter V.

Summary

In this study, evaluation instruments were developed on objectives which met one or more of three criteria (considered by experts as high priority, or related to disease entities causing either high morbidity or mortality). The instruments, written and observation, were the subject of reliability and validity studies in which they were administered to a sample of eight CTN students as pre- and post-tests.

A classical research design of randomly selected sample of CTN's with a control group of RN's was considered appropriate for utilization of the instruments in on-site evaluation. To enhance validity, the proportional random sampling design included stratification of the CTN's into recent/non-recent, and of nursing stations into large and small. To assess the influence of selected extraneous antecedent variates of nurses, analysis was done of data from a sample of thirty-two students.

The results and discussion of the data analyses arising from investigations described in this chapter are presented, following, in Chapter IV.

CHAPTER IV

RESULTS AND INTERPRETATION OF DATA ANALYSES

Introduction

In this study, as described in Chapter III, there were two main areas of investigation, the development of measurement instruments of known reliability and validity, and the determination of a research and sampling design suitable for the incorporation of the instruments into a field evaluation package. The results of data analyses are herein reported in terms of these two areas, although within each section there were various data sources. A probability level of .05 for a one-tailed test was used as the level of significance in analyses of correlations and of pre-post differences; alpha otherwise equals .05, two-tailed.

Measurement Instruments

Areas of Objectives Selected

There were three criteria by which objectives were selected for inclusion in the evaluation: validation by CTN educators, and relevent to disease categories causing either high morbidity or mortality (cf. pp. 32-35). In Table 2 are shown the content areas of the bank of objectives which met the criteria. All content areas on which items were based met at least one criterion, a majority (58.6%) meeting more than one. Items were developed to assess performance on all content areas

Table 2.

Content Areas Selected for Evaluation, by Criteria

Criterion 1. Objectives Validated by Experts as High Priority	Criterion 2. Diseases Causing High Morbidity[a]			Criterion 3. Disease Categories Causing High Mortality[b]	
	code		frequency		mean no. of deaths
acute gastroenteritis & dehydration	*186	special conditions & examinations without sickness	5112	*accidents	76.0
genitourinary infections				*circulatory diseases	35.5
*partial history & physical	148	other diseases of skin & subcutaneous tissue	4339	*respiratory diseases	29.5
meningitis				neoplasms	29.0
venereal disease	*092	acute upper respiratory infection	3949	perinatal causes	19.5
iron deficiency anemia					
*otitis media	*178	laceration, open wound, superficial injury, contusion, crushing with intact skin surface	3686		
*congestive heart failure					
acute abdomen					
*obstetrical history					
*obstetrical physical	*003	tuberculosis	3175		
*pap smear	097	hypertrophy of tonsils & adenoids	2936		
	*073	otitis media	2533		
	146	infections of skin & subcutaneous tissue	1613		
	*095	bronchitis & emphysema	1570		
	*100	other diseases of upper respiratory tract	1507		
	*093	influenza	1128		
	104	diseases of teeth & supporting structures	1108		

*--Subject areas meeting more than one criterion.

a--Northern Region, 1973
b--Northern Region, 1972 and 1973

meeting a criterion.

Analysis of N.W.T. morbidity data further indicated that 52.4% of patients seen were adults, and therefore approximately one-half of the evaluation content was mede referable to adults, the remainder to infants and children.

Written Examinations

The alpha reliability coefficient over all written tests was .89, an indication that both written tests were internally consistent. Inter-correlations of pre- and post-test scores on the written examinations are provided in Table 3. Values underlined can be regarded as reliability coefficients, since the same instrument was used (even though treatment intervened). Those values in parentheses are concurrent estimates of validity.

Table 3.

Correlations[a] Among Written Examinations

			General		Obstetrics	
			Pre (RN's)	Post (CTN's)	Pre (RN's)	Post (CTN's)
General	Pre	(RN's)	1.00			
	Post	(CTN's)	.94	1.00		
Obstetrics	Pre	(RN's)	(.46)	.66	1.00	
	Post	(CTN's)	.67	(.81)	.81	1.00

Note. Values in parentheses are estimates of validity, those underlined of reliability.
a--For n=8, correlations must exceed .62 to be statistically significant.

In Table 3, the only correlation which is not significant is the concurrent validity estimate of the written examinations for RN's (0.46), but the concurrent validity estimate for the same tests on CTN's is 0.81. This disparity may be partially accounted for by the fact that the material in the examinations is intended to measure expertise at the CTN level. Assuming RN's lack such knowledge, it is not surprising that the correlation of their scores on the two tests is statistically regarded as random. As will be seen, this interpretation is supported by the analysis for construct validity. In Table 3, all other reliability and validity estimates are sufficiently high to be acceptable.

Reasonable construct validity of the written tests can be assumed since scores of the CTN's (post-tests) were significantly greater than those of RN's (pre-tests), as shown in Table 4.

Table 4.

Analysis of Pre-Post (RN-CTN) Mean Differences[a] on Written Examinations

Examination		Mean	Source of Variation	sum of squares	df	mean squares	f
General	Pre-test	56.38					
	Post-test	90.00					
			between people	4469.94	7	638.56	
			within people	4678.50	8	584.81	
			treatments	4522.56	1	4522.56	203.02
			residual	155.94	7	22.28	
			total	9148.44	15		
Obstetrics	Pre-test	41.50					
	Post-test	78.13					
			between people	1291.94	7	184.56	
			within people	5504.50	8	688.06	
			treatments	5365.56	1	5365.56	270.33
			residual	138.94	7	19.85	
			total	6796.44	15		

a--critical value of $_{.90}F_{1,7} = 3.59$ (F-value for one-tailed test, $\alpha = .05$)

Observation Instruments

The inter-judge reliability and inter-professional validity estimates for the observations of nurses (pre-tests) and CTN's (post-tests) are given in Table 5, where the judges' professional levels are abbreviated as: medical doctor, MD; nurse practitioner, NP; nurse, RN; and nurse-educator, CTN-NE. The figures one (1) and two (2) associated with each identify the two individual members of each group of professionals utilized as raters. The observations included the live marking of the suturing, and judging from recordings of the other performances. As inadequate recordings were made of two students' pre-tests on the paediatric history, the sample size was reduced to six for the instrument for nurses. For the most part, the correlations are statistically significant, indicating reliability and validity (as previously defined, cf. pp. 40-41) as being at acceptable levels.

Some difficulty in measuring the suturing skills of RN's (pre-tests) with the instrument developed is evident from Table 5. The relatively low validity may in part be due to certain difficulties encountered at pre-testing, namely:

a) more than one physician was needed to complete the scoring;

b) lack of rigor on the part of some raters in marking according to the criteria specified; and

c) a philosophical disagreement by one physician with the format of the objectives and the instrument for evaluating performance of them. This may well have contributed to the unreliability noted in b).

Conversely, the instrument may in fact have lower validity for evaluating the suturing skills of nurses than of CTN's, the reason being similar to

Table 5.

Correlations of Judges on Observation Examinations: Inter-judge Reliability[a] and Inter-professional Validity Estimates[b]

Physical Examination

Group	Judge	MD1	MD2	NP1	NP2	RN1	RN2	CTN-NE
NURSES	MD1	1.00						
	MD2	.89	1.00					
	NP1	.96	.91	1.00				
	NP2	.96	.96	.95	1.00			
	RN1	.93	.96	.90	.94	1.00		
	RN2	.85	.90	.92	.88	.87	1.00	
	CTN-NE	.92	.84	.96	.88	.84	.94	1.00
CONTROLS	MD1	1.00						
	MD2	.89	1.00					
	NP1	.88	.88	1.00				
	NP2	.91	.93	.75	1.00			
	RN1	.79	.77	.55*	.85	1.00		
	RN2	.78	.93	.77	.81	.77	1.00	
	CTN-NE	.92	.95	.84	.94	.86	.92	1.00

Adult History

Group	Judge	MD1	MD2	NP1	NP2	RN1	RN2	CTN-NE
NURSES	MD1	1.00						
	MD2	.96	1.00					
	NP1	.44*	.44*	1.00				
	NP2	.88	.82	.51*	1.00			
	RN1	.96	.93	.57*	.91	1.00		
	RN2	.97	.95	.53*	.90	.99	1.00	
	CTN-NE	.97	.94	.36*	.90	.92	.94	1.00
CONTROLS	MD1	1.00						
	MD2	.89	1.00					
	NP1	.94	.87	1.00				
	NP2	.94	.86	.89	1.00			
	RN1	.95	.88	.93	.99	1.00		
	RN2	.93	.86	.83	.97	.96	1.00	
	CTN-NE	.98	.86	.89	.96	.95	.95	1.00

Paediatric History

Group	Judge	MD1	MD2	NP1	NP2	RN1	RN2	CTN-NE
NURSES	MD1	1.00						
	MD2	.97	1.00					
	NP1	.92	.98	1.00				
	NP2	.70*	.79	.80	1.00			
	RN1	.73	.82	.81	.69*	1.00		
	RN2	.91	.91	.87	.86	.72*	1.00	
	CTN-NE	.91	.98	.97	.83	.82	.87	1.00
CONTROLS	MD1	1.00						
	MD2	.94	1.00					
	NP1	.99	.97	1.00				
	NP2	.98	.97	.98	1.00			
	RN1	.95	.97	.94	.97	1.00		
	RN2	.95	.84	.93	.91	.81	1.00	
	CTN-NE	.99	.93	.98	.96	.93	.95	1.00

Suturing

Group	Judge	MD	NP	RN	CTN-NE
NURSES	MD	1.00			
	NP	.29*	1.00		
	RN	.58*	.61*	1.00	
	CTN-NE	.69	.64	.79	1.00
CONTROLS	MD	1.00			
	NP	.80	1.00		
	RN	.77	.73	1.00	
	CTN-NE	.95	.63	.78	1.00

a--Reliability estimates are underlined.
*--No significant correlation.

that provided in the discussion of the concurrent validity estimate of written tests (cf. p. 50).

In addition, the reader's attention is drawn in Table 5 to the low reliability and validity coefficients of the adult history instrument when used by one nurse practitioner to evaluate nurses. Since the corresponding coefficients are at acceptable levels on the paediatric history, a very similar form, a reasonable interpretation of these findings would be lack of rigour by the judge during marking of the pre-tests. However, with only two nurse practitioners actually having used the adult history instrument, the reason for the low correlations must remain largely speculative. With the exceptions noted, the reliability and validity coefficients indicate that the instruments can be considered reliable and valid.

The average validities and reliabilities, derived using the Fisher's Z-transformation of the original coefficients (Glass & Stanley, 1970, p. 534), are shown in Table 6. The inter-judge reliability estimates, previously reported, are reproduced in Table 6 to facilitate comparison with the mean inter-professional validity estimates. The latter values were obtained by averaging the correlations of the judges in a professional level with all other judges. In section 5 of this table are the estimates of reliability and validity of all observation instruments when used by each professional level to separately evaluate nurses and CTN's; in section 6 are corresponding estimates when used to evaluate RN's and CTN's. The reader's attention is drawn to the fact that the acceptability of the reliability and validity estimates is not confined to the use of the instruments by any one of the groups of judges. Rather, the results indicate that nurses, physicians, nurse

Table 6.

Mean Inter-judge Reliability and Inter-professional Validity Estimates for Observation Instruments

Observation Instrument		Professional Level of Evaluators: A. Physicians		B. Nurse Practitioners		C. Nurses		D. CTN Nurse-Educator		E. All judges		F. All judges
		Subjects Evaluated: RN's	CTN's	RN's	CTN's	RN's	CTN's	RN's	CTN's	RN's	CTN's	RN's & CTN's
1. Physical Examination	Rel.	.892	.887	.960	.750	.866	.765			.917	.814	.875
	Val.	.929	.889	.933	.860	.900	.825	.908	.913	.923	.870	.900
2. Adult History	Rel.	.955	.893	.506*	.894	.993	.960			.938	.924	.931
	Val.	.907	.920	.740	.936	.908	.940	.911	.944	.856	.910	.886
3. Paediatric History	Rel.	.966	.945	.802	.984	.721*	.812			.878	.942	.915
	Val.	.902	.966	.879	.966	.842	.941	.921	.965	.885	.960	.932
4. Suturing	Val.	.544*	.866	.530*	.727	.670	.762	.716	.835	.623	.805	.726
5. All instruments	Rel.	.949	.915	.835	.920	.926	.877			.914	.905	.909
	Val	.860	.920	.818	.903	.857	.893	.881	.926	.856	.910	.886
6. All instruments (RN's & CTN's combined)	Rel.	.934		.885		.905				.910		.910
	Val.	.895		.867		.873		.906		.886		.886

Note. The abbreviations "Rel." and "Val." are used for inter-judge reliability and inter-professional validity.
*--not statistically significant.

practitioners, or CTN nurse-educators (if this investigator is representative of the last) can dependably perform as evaluators using these particular observation instruments. Further, in section E. of Table 6 are the reliability and validity estimates for the instruments when used by any of the professional levels to evaluate separate groups of nurses or CTN's, and in section F. are the corresponding estimates of the inter-judge reliabilities and inter-professional validities of the instruments for the evaluation of nurses and CTN's--when used by the professional levels represented in validation studies.

The validity data for the physician's live observation of physical examinations, not included in the previous tables, are shown in Table 7. The coefficients are all at acceptable levels, and comparable to coefficients in video-taped observations. This similarity provides support for recommendation of the use of the instruments in live observation for field evaluation.

As discussed previously (cf. pp. 17-18), several formulae may be employed in estimating reliability, appropriate applications for which are as follow in Figure 2.

Standardization of raw scores	Reliability coefficient	
	Any one judge or occasion	All judges or occasions
nil	unadjusted r_1	unadjusted r_k
to same mean	adjusted r_1	adjusted r_k
to same mean and variance	r_{xy}	α

Figure 2. Applications of reliability coefficients.

Table 7.

Validity and Reliability[a] Estimates of Physical Examination Instrument: Live and Video-taped Observations.

Observation	Nurses	CTN's	Mean of Nurses and CTN's
Physician (1) Live & Physician (1) VTR	.955	.850	.918
Physician (1) Live & Physician (2) VTR	.913	.766	.854
Physician (1) VTR & Physician (2) VTR[a]	.892	.887	.890
Mean of physician (1) Live and all VTR	.909	.749	.865
Mean of physician (1) VTR and all VTR	.927	.874	.904
Mean of all Live and VTR	.913	.855	.888

Note. All are statistically significant

VTR refers to judging from video-taped recordings.

a--inter-judge reliability estimates (cf. Table 5.)

Further reliability estimates for the observation instruments are given in Table 8, the applications of which were indicated in Figure 2. The reliability estimates increase with the degree of standardization of raw scores (there being no reliability estimate less than .81 with standardization to the same mean and variance). Standardization is particularly necessary for the suturing instrument, acceptable reliability in evaluating RN's being achieved if the linear relationship alone is of importance.

Judges were very consistent in their linear ratings (as indicated by the intra- and inter-professional correlations of Tables 5 and 6), but differ in means assigned (cf. F-values, Table 8). The elimination of these sources of inconsistency was accomplished by standardizing all scores for an instrument's applications to the (1) grand mean (pre- and post-tests), and (2) total variance of the scores which remained after extraction of the variability due to there being several judges. The resulting variability (standard deviation) for each instrument, also reported in Table 8, was arrived at by the formula:

$$S.D.=\left(\frac{\sum_{i=1}^{n_j}\sum_{j=1}^{J}\sum_{k=1}^{K}(X_{ijk}-\bar{X}...)^2-\sum_{j=1}^{J}n_j\sum_{k=1}^{K}(\bar{X}._{jk}-\bar{X}._{j}.)^2-\sum_{j=1}^{J}\sum_{i=1}^{n_j}\sum_{k=1}^{K}(X_{ijk}-\bar{X}_{ij}.)^2}{(K-1)(n_1+n_2)-JK+2}\right)^{.5}$$

where X_{ijk} represents the ith subject on the jth test and scored by the kth judge. This formula includes in the variability that variance which is: between people, residual, and of the pre- and post-test means about the grand mean of pre- and post-tests. Excluded from the variability is that of the judges around the means of pre- or post-tests, and within person variability on the pre- and post-tests.

Table 8.

Central Tendencies and Dispersions of Raw Scores, and Reliability Estimates: Observation Examinations

Instrument	Grand mean of pre- and post-tests	Standard deviation of pre- and post-tests	Test	F^a	Unadjusted		Adjusted		Linear Relationship	
					r_1	r_k	r_1	r_k	r_{xy}^b	α
Physical Examination	104.44	39.58	Pre	8.68	.81	.97	.87	.98	.92	.98
			Post	14.51	.62	.93	.80	.97	.81	.97
Suturing	34.36	9.41	Pre	1.57*	-.06	-.27	-.04	-.19	-	.85
			Post	3.95*	.49	.79	.58	.85	-	.92
Adult History	64.16	20.65	Pre	8.02	.63	.92	.77	.96	.94	.96
			Post	29.39	.67	.93	.91	.99	.92	.99
Paediatric History	52.26	19.16	Pre	9.18	.67	.93	.81	.97	.88	.97
			Post	8.08	.88	.98	.94	.99	.94	.99

a--The F-values are those resulting from analysis of variance among judges (mean differences within the pre-test and post-test). The critical value of $_{.95}F_{1,7} = 5.59$, and of $_{.95}F_{1,4} = 7.71$, the latter being applicable to the paediatric history.

b--Mean interjudge reliability (cf. Table 9).

*--Not statistically significant.

Note: The grand mean and standard deviation of the written General and Paediatrics examination were 73.19 and 23.92, and of the Obstetrics and Newborn, 59.81 and 20.62.

A test of construct validity was of mean differences on pre- and post-testing. In Table 9 are set out, for each judge on all observation examinations, the F-values of both the raw scores and corresponding scores after standardization to the means and standard deviations in Table 8 (cf. p. 58). One could conclude that since all mean differences are statistically significant, the nurses (pre-tests) have less ability than when they are CTN's (post-tests). These results give evidence of construct validity, in that the instruments reflected the expected change in student behavior in response to the intervention of the CTN course.

The t-tests (Welch's prime adjustments) on analysis of pre- and post-test scores as measures of independent groups (results of which are given in Table 10) were also all significant. In addition to the values shown, those for the written General and Paediatrics (3.70) and Obstetrics and the Newborn (7.25) were also significant. These mean differences may be interpreted as indicating that the construct validity of the instruments is great enough that they will discriminate between CTN's and an independent control group of RN's in field evaluation.

Incorporation of Instruments into Field Design

For analyses of the influence of antecedent variates on the achievement of previous University of Alberta students (cf. pp. 44-45), the subtotal pre-tests included thirty-two students on the suturing and physical examinations. One class of eight students was excluded from the sample for total pre-tests, because of the lack of common content on adult history and written examinations. The sample size for the history, written, and total pre-tests was thus twenty-four; for total post-tests, thirty-two.

Table 9.

F-values of Pre-Post (RN-CTN) Mean Differences in Raw and Standardized Scores: Observation Instruments

		Physical Examination[a]	Suturing[a]	Adult History[a]	Paediatric History[b]
RAW SCORES	MD1	197.83	58.98	170.71	24.92
	MD2	142.75		142.91	33.21
	NP1	170.77	10.84	220.40	39.75
	NP2	163.60		130.88	39.87
	RN1	248.42	55.42	179.17	35.19
	RN2	215.74		126.52	53.39
	CTN-NE	446.21	54.15	200.95	42.25
STANDARDIZED	MD1	194.33	59.53	175.12	24.26
	MD2	139.62		140.63	33.03
	NP1	171.18	10.60	227.15	38.85
	NP2	169.18		125.58	38.98
	RN1	254.47	54.52	175.51	37.43
	RN2	211.29		120.87	51.45
	CTN-NE	434.14	50.66	198.03	41.38

Note. All are statistically significant.

a-- n=8, the critical value of $_{.90}F_{1,7} = 3.59$ (one-tailed test, $\alpha = .05$)

b-- n=6, the critical value of $_{.90}F_{1,5} = 4.06$ (one-tailed test, $\alpha = .05$)

Table 10.

T-tests (Welch's Prime Adjustments) of Mean Differences on Observation Instruments Using Pre- and Post-tests as Measures of Independent Groups

Judge	Instrument			
	Physical Examination	Suturing	Adult History	Paediatric History[a]
MD 1	10.02	5.93	12.00	6.42
MD 2	8.02		10.49	6.94
NP 1	9.01	3.22	11.17	7.57
NP 2	8.57		10.97	6.51
RN 1	10.67	6.46	12.04	6.22
RN 2	11.19		10.22	7.91
CTN-NE	10.77	6.44	14.26	7.78

Note. All are significant, the critical value of $_{.95}t_{14} = 1.761$, and $_{.95}t_{12} = 1.782$.

a--n=14, for all other, n=16.

The results of the analysis of variance approach (shown in Table 11) indicate that those nurses (RN's) who have a year-long midwifery course can be expected to score higher on a CTN written obstetrics examination than nurses without midwifery. The sample for the obstetrics pre-test included twenty nurses with no midwifery training, and four with a one year midwifery course. Of the post-test sample, twenty-five CTN's had no midwifery training, two had a six-month course, and five a full year's training. For analysis, this larger sample was trichotomously stratified, results indicating that CTN's with a previous full year's midwifery training can be expected to achieve significantly higher scores than CTN's without midwifery.

Analysis of variance additionally indicated that nurses who have had at least one year's experience in nursing stations will suture more proficiently than nurses with less experience in such a setting. It should be noted that there was no systematic difference in suturing skills between CTN's with and without nursing station experience (of at least one year). With these exceptions, if the tests used at the University of Alberta are representative of the evaluation package, overall performance of RN's and CTN's as indicated by analysis of variance, will not be significantly affected by their: midwifery or other educational background, age, or length of experience in nursing stations.

Step-wise regression analysis for the influence of the antecedent variates as predictors of achievement on the composite tests gave some results which conflicted with the analysis of variance approach for RN's. On total pre-test, 17.32% of the variance in test scores was accounted for by length of nursing station experience, and 20.26% of the remaining variance

Table 11.

Influence of Antecedent Variates:
Probabilities Associated with F-values under Null Hypothesis

Examination		Antecedent Variate			
		midwifery	education	experience	age
Physical exam, performance	pre	.25	.46	.23	.74
	post	.87	.30	.05	.98
Physical exam, technique	pre	.11	.52	.11	.97
	post	.19	.10	.14	.43
Suturing	pre	.92	.48	.04*	.63
	post	.85	.98	.19	.10
Adult history	pre[a]	.81	.13	.97	.81
	post	.21	.38	.25	.90
Written: General	pre[a]	.06	.84	.46	.99
	post	.29	.58	.75	.68
Paediatrics	pre[a]	.27	.90	.08	.90
	post	.22	.54	.63	.08
Obstetrics	pre[a]	.01*	.61	.47	.80
	post	.03*	.65	.41	.43
Subtotal of tests	pre	.27	.74	.05*	.76
Total	pre[a]	.07	.63	.33	.77
	post	.57	.34	.92	.86

a--n=24, otherwise n=32.

*--statistically significant

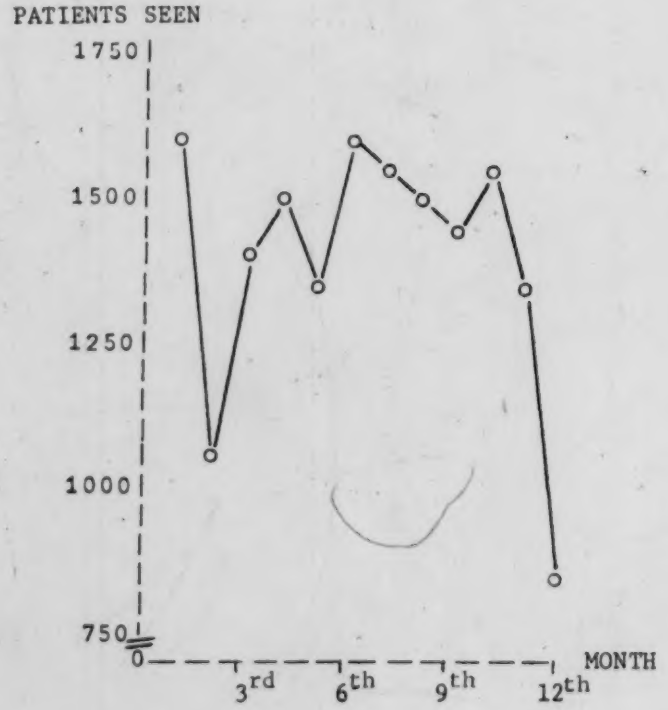
by midwifery background. These were the only significant results. This would indicate that dichotomizing and trichotomizing continuous data for experience and midwifery training had concealed identifiable sources of variance in RN results. The step-wise regression analysis supported the findings of the analysis of variance model on post-tests--none of the antecedent variates was a significant predictor of total evaluation scores for CTN's. In future evaluations, therefore, it would be necessary to use these experience and midwifery covariates, as no reasonable categorization could be obtained for stratified sampling.

Sampling

Examination of the population size of the communities in which nursing stations are located indicated there are 43 with populations of at least five hundred, and 29 smaller communities (cf. Appendix H). The sampling design recommended is based on proportional random sampling of this population of nursing stations in which approximately 60% provide service to communities with populations of at least five hundred. Patient care observations are necessary to complete the evaluation of a subject with the instruments. As a .59 correlation between the number of patients seen in a nursing station and the size of the community had been found, this sampling of additional large communities would result in an increased likelihood of the observer completing the recommended observations.

In relation to maximizing the likelihood of the required patients presenting during the on-site visit, N.W.T. and southern nursing stations' frequency of patient visits were analyzed for seasonal fluctuations. As illustrated in Figure 3, in Northern Region there is a tendency toward

N.W.T. Nursing Stations



Southern Nursing Stations

Figure 3. Total patient visits in 1973 to all N.W.T. and five southern nursing stations, by month.

increased visits in the fall and early winter; in the south, more consistency is seen across months with the exception of the decrease in November and December. On-site evaluation taking place from mid-August to mid-December, beginning with southern nursing stations, would thus yield the largest number of patient visits on which the nurses' practice would be evaluated.

It had been a request of Medical Services Branch that the field evaluation design permit inferences to be made as to the comparative retention in practice of CTN course material by recent and non-recent CTN's. Accordingly, in the design recommended, illustrated in Figure 4, sampling is also proportional to the estimated percentages of RN's, recent, and non-recent CTN's staffing nursing stations.

Nursing stations	Community population		
	≥ 500	< 500	
with recent CTN's	randomly select 2	randomly select 1	3
with non-recent CTN's	randomly select 2	randomly select 2	4
with RN's	randomly select 4	randomly select 3	7
	8	6	14

Figure 4. Sampling design for field evaluation (indicating proportional random sampling of nursing stations stratified by size, and CTN's by time since course completion).

The sampling design was based on the premise that 50% of nursing station staff will be CTN's, with fewer recent than non-recent graduates in service at the time of the field evaluation. It was the opinion of Medical Services Branch personnel that approximately 50% of the CTN's in nursing stations are non-recent graduates, some 150 having completed a course by December, 1974 (Medical Services Branch, 1974). From data available for Northern Region (presented in Table 15), it would appear that the attrition rate for non-recent CTN's may be lower than estimated. This sample may not have been representative, however, because of the influence of such factors as geographical location and/or the high proportion of CTN's posted to nursing stations in the N.W.T. (vis-à-vis other types of facilities such as health centres and hospitals).

The factors leading to the size of the sample of nursing stations include:

1) the lack of data to support fewer than eight subjects in either the experimental or control groups; and
2) Medical Services Branch's pattern of staffing the majority of nursing stations with more than one professional nurse.

Therefore, random sampling of the nursing stations should be without replacement, reasonably assuring assessment of sixteen subjects (the inferred independent sample size used in this study); potentially, however, twenty-eight subjects could be assessed and this would increase the power of the evaluation.

Summary

Written and observation instruments were developed on the twenty-nine areas of objectives which met the criteria for inclusion in the

Table 12.

Retention Rates and Central Tendencies of Length of Service (in months) of CTN's Sponsored by Northern Region, MSB

Finishing date, CTN course	Number in sample	% still employed by MSB	LOS since course: central tendencies			Mean total LOS	Mean LOS prior to CTN course	Mean LOS prior to CTN course for year[b]
			Mean	Median	Mode			
April 30, 1972	8	62.5	23.6	21.5	37	41	14	15.5
Feb. 14, 1973	6	50.0	19.5	26.5	27	42	18	
July 7, 1973	7	85.7	21.2	21.9	22	34	9	14.5
Feb. 14, 1974	8	75.0	12.7	14.8	15	37	20	
July 14, 1974	8	75.0	9.1	9.9	10	19	6	12.0
Feb. 14, 1975	7	85.7	3.0	3.0	33	43	18[a]	
July 1975	9						17	

Note. Finishing dates estimated from University of Alberta information, with calculations to May 16, 1975. The abbreviations MSB and LOS refer to Medical Services Branch, and length of service.

a--this class included one student who had worked 122 months with Northern Region: if her data is included, it brings the mean to 36 months.

b--year refers to MSB's fiscal year.

evaluation. Studies undertaken of the instruments indicated that all had acceptable levels of both reliability and validity. Further, results indicate that with few exceptions, the observation instruments will yield similar results when scored by physicians, nurses, nurse practitioners, or CTN nurse-educators, provided scores are standardized to a given mean and variance.

Since the data available regarding the patterns of staffing nursing stations in terms of RN's and recent/non-recent CTN's were inconclusive, the proportional stratified sampling design was developed on assumptions which can be tested prior to a field evaluation. Two antecedent variates were found to influence RNs' performance, the influence of which can be controlled by the use of covariance in data analysis. These variables were midwifery background and length of experience in nursing stations.

In the final chapter, a summary of this study and recommendations arising from it are presented.

CHAPTER V

SUMMARY AND RECOMMENDATIONS

Summary

The literature indicates that in the widespread development of the extended role of the nurse in primary health care, considerable investigation has taken place of the degree of acceptance, job satisfaction, and legal constraints associated with the role. Evaluation of performance of primary care nurse practitioners has been largely descriptive and/or comparative with physicians' performance.

Clinical Training for Nurses (CTN) courses are offered at several Canadian universities in response to the perceived need (Kergin, 1970) for the upgrading of clinical skills of nurses employed in Medical Services Branch's isolated nursing stations. An important facet of the CTN courses is the rigourously and co-operatively developed bank of behavioral objectives which forms the core curriculum for the courses, and hence the knowledge and skills expected of all CTN's. Although each university is charged with the responsibility of evaluating CTN's upon course completion, no objective evaluation has been done of performance in practice.

The purposes of this study were two-fold:

(1) to develop instruments of known reliability and validity for the measurement of knowledge and skills of CTN's; and

(2) to incorporate the measurement instruments into a research design for field evaluation.

The focal instruments of this project were written examinations (Obstetrics and the Newborn, and General and Paediatrics), and instruments for evaluation by observation of suturing, physical examination, and history-taking for adult and paediatric patients. All instruments were based on the content of the bank of objectives.

The sample for validation studies was a class of eight University of Alberta CTN students who were considered RN's on pre-testing, and CTN's on post-testing. Each observation examination was judged by two physicians, two nurses, two nurse practitioners, and one CTN nurse-educator, the first three being considered representative of their professional groups because of the nineteen individuals involved. Evaluation of suturing skills was done during performance, but video and audio recordings were made of the simulated patients' physical examinattions and interviews to allow independent marking by the judges without the limitation of uncontrolled observer effect.

A probability level of .05 (two-tailed) was used throughout except for analyses of correlations and comparisons of pre- and post-test data. For all examinations, reliability estimates were obtained by the use of the alpha coefficient. In addition, inter-judge (intra-professional) reliability estimates were derived for the observation instruments. Empirical estimation of concurrent and construct validity (including the use of inter-professional correlations on observations as a measure of concurrent, and mean differences from pre- to post-testing as construct validity) yielded highly acceptable levels of validity for all instruments.

Although results of validation studies indicated that the instruments developed could provide valid data when used by any of the professional

groups from which the judges were drawn, certain threats to generalizability are inherent in a field evaluation in which the instruments would be utilized. Data from a sample of thirty-two CTN students, analyzed for the influence of antecedent variates, indicated that previous midwifery training and length of nursing station experience affect the performance of nurses without the CTN course, but not CTN's. This source of internal invalidity, due to practical constraints, would necessarily be controlled by the use of covariates in data analysis. The sampling design for field evaluation does allow comparison of recent and non-recent CTN's, as well as CTN's with a control group of RN's. It also controls for the effect of the size of the community in which the nursing station is situated.

The limitations of this study include a small sample size, several assumptions regarding the representativeness of the sample and judges, and uncertainty regarding the accuracy of morbidity and staffing data. However, if the recommendations which follow are heeded and the instruments used as described in the accompanying training manual, the purposes of this study will have been fulfilled. The measurement instruments, of known reliability and validity, are capable of identifying differences in behavior due to the CTN course in the educational setting, and should do so in the field if changes remain after the CTN's return to practice.

Recommendations

Recommendations arising out of this study fall into two areas: those specifically dealing with design implementation in field evaluation, and those of a more general nature.

Recommendations for Field Evaluation

The following recommendations for the implementation of the instruments and design in an on-site evaluation are based on findings of this study, and the investigator's observations of and discussions with participants in the project. In the interest of cogency, the recommendations are divided into several subject areas.

Instruments. Due to the highly acceptable levels of reliability and validity, it is recommended that the instruments which were a focus of this study be utilized in field evaluation of CTN's as described in the Instruction Manual (cf. Appendix G).

Sampling. To control for the effects of community size as well as the inter-active effects of time and the CTN course, it is suggested that the sampling design provided on page sixty-six be used. In view of rapid population shifts in some areas, the stratification of nursing stations should be corrected to the most current community population figures available at the time of the on-site assessment (although Appendix H lists the nursing stations stratified by 1973 community population).

As validation studies have indicated that the instruments composed in this study discriminate between sample sizes as small as eight, it is recommended that a minimum of eight CTN's with an equivalent control group of RN's be evaluated. Immediately prior to the field evaluation, information as to the staffing of all nursing stations (in terms of recent and non-recent CTN's and RN's) should be obtained and the nursing stations stratified for the staffing variable prior to random selection. The total number of nursing stations to be randomly sampled with replacement is reflected in the sampling design. If a station has more than one

nurse (RN or CTN), two nurses should be evaluated during the on-site visit. If more than two nurses are resident, the evaluation should be of that nurse for whom the station was originally selected, and one other, also randomly chosen. Thus, a potential total of twenty-eight RN's and CTN's may be evaluated within the time period later discussed.

As pre- and post-test sensitization can reduce external validity, the CTN's who participated in validation studies can no longer be considered representative of the target population, and therefore should be excluded from the field evaluation sample. A major threat to internal validity is the possibility of change in the observer during the duration of the evaluation. To avoid systematic effect of this instrumentation (cf. p. 27), the order in which CTN's and RN's are evaluated should be randomized.

Evaluator. The validation studies have indicated that any one of the following professional categories is an acceptable group from which an evaluator can be chosen: physicians, nurses, nurse practitioners, (including CTN's), or CTN nurse-educators. This is evidenced by the correlations between two physicians' ratings being no higher than the physicians' ratings with the other judges. This recommendation on the professional qualifications of the evaluator is made with confidence as:

a) the pattern of inter-professional agreement was as high or higher than the intra-medical agreement across the four observation tests; and

b) this pattern was unlikely to be a function of particular individuals because of the number of judges involved in this study (cf. pp. 41-42).

The exception to this recommendation is the CTN nurse-educator, as no peers were available to act as judges. It may well be that intimate acquaintanceship with the project may have lessened her representativeness of CTN nurse-educators.

Keeping in mind that the professional level of the evaluator (within the groups represented as judges in validation studies) is not a crucial factor in obtaining accurate data, consideration in selection must therefore be given to budgetary constraints and the availability of personnel. As all professional categories use the observation instruments at an acceptable level of validity, the number of evaluators should be a function of: (1) the expediency of completion of data collection, and (2) the cost and ease of selection and training. Taking these factors into account, it is suggested that the most cost-effective method would be to have one person do the on-site assessments.

In choosing the evaluator, personal qualities are of considerable importance if the acceptability of the evaluation is to be maximized, observer effect minimized, and accurate data obtained. The following recommendations regarding personal qualifications are based on experience during validation studies, and discussions with nursing personnel and CTN's.

1) It is recommended that the evaluator be adaptable, acceptable, and non-threatening. As the observer will live with, or in near proximity to the subjects, he should be capable of accommodating readily to nursing station environment. Discussions with personnel of the Universities of Alberta, Toronto, and McMaster who have been involved in observer evaluation, support the position that adjustment problems are less severe if a nurse (rather than a physician or someone with no health care background) acts as

the evaluator. Further, to avoid any semblance of personal evaluation, it would be advantageous if the evaluator did not have line responsibility for the subjects.

2) Additionally, the evaluator should be capable of objectivity and attentiveness to detail. Experience during this study indicated the need for uninterrupted concentration on the task at hand, and ability and willingness to focus on finite details of specific observable behaviors. Without concentration, precision, and consistency on the part of the observer, the subjects may not be correctly credited.

3) Further, the evaluator should be accepting of the philosophy of the evaluation, and the bank of objectives on which it was based. During the development of both the bank of objectives and the evaluation instruments, it was apparent that some health professionals adhere to philosophies which do not allow a commitment either to specifically spelling out the knowledge and skills which a CTN requires, or to assessing nurses on a sample of such a universe. As such attitudes are in direct conflict with the rigorous objectivity demanded of the observer by the instruments, they must be either scrupulously avoided in or suppressed by the evaluator. Additionally, the evaluator's objectivity should extend to the CTN courses, in that he should have no vested interest in whether or not the courses are continued.

Length and time of visits. Based on statistical analysis of N.W.T. data, daybooks from southern stations, and the consideration of observer effect and time required to collect the data, it is recommended that each visit be five days in length ("days" being those in which a nursing station is open for provision of regular services). It is estimated that if the

visits take place during the months which are recommended, following, most nursing stations will have the necessary patients present on which the nurse is to be evaluated during the five day period.

From analysis of seasonal fluctuations in frequencies of patient visits, it is recommended that on-site evaluation take place from mid-August to mid-December, beginning with southern nursing stations. This eighteen-week period would allow coverage of fourteen stations, inclusive of the observer's "days off" and travel time.

Staff preparation. As it is anticipated that all subjects will have some anxiety regarding the evaluation, it is recommended that a full explanation of the evaluation be forwarded to them well in advance of the visit. This measure should help to reduce the observer effect, as well as the anxiety created by an "unknown" situation. All medical and nursing directors in the regions concerned should also be well informed in advance, to ensure their interest and co-operation, and avoid conflicts in schedules.

Data analysis. The following recommendations for statistical analysis and interpretation of data are based on the data from validation studies as well as the sample of previous CTN students at the University of Alberta.

If more than one evaluation of a subject is done with a particular evaluation form (e.g., two lacerations presented and thus two measures of suturing were obtained), the mean of each item for the subject should be used in subsequent analyses.

Analysis of covariance should be done on the scores for each test as well as the total of tests, wherein:

1) months of experience in nursing stations as an RN is one covariate,
2) whether or not the subject is a certified midwife is the other covariate, and

3) levels for distinguishing between experimental and control groups are: a) firstly; recent CTN, non-recent CTN, and RN, and

b) secondly; CTN and RN.

The use of the covariates will allow interpretations to be made independent of RN experience or midwifery training. That is, if the CTN course does make a difference in nurses' practice, it can be stated that the difference is true when RN's and CTN's are made statistically equivalent for those antecedent variables found to influence achievement on CTN examinations.

Dependent t-tests should be performed on the following two criteria:

1) the elicited scores (E) on history and physical, compared to

2) the corresponding record review scores (R) for history and physical.

If a significant difference arises, the unvalidated record review should be suspect, and subsequently omitted from analyses in which RN's and CTN's are compared.

As the size of the target population will be known in the evaluative field study, the standard error estimate can be modified for a finite population by the factor $(1 - \frac{\text{\# of RN's and CTN's in sample}}{\text{\# of RN's and CTN's in population}})^{.5}$.

Interpretations of the findings of the field evaluation must be limited to "There is or there is not an improvement in the quality of practice (as indicated by the sample criteria on evaluation forms) of nurses who have taken the CTN course." If interpretation is required as to how well RN's or CTN's perform, the evaluator's ratings should be standardized to the means and standard deviations established in the validation study (cf. Table 8, p. 58). This is necessary as, although the intercorrelations among judges were high, differences in means and standard deviations are a source of inconsistency not revealed by the correlation coefficients.

Follow-up. It is anticipated that the subjects will wish to have feedback on their performance, but it must be made clear that no individual's score will be calculated, nor will any system of pass/fail be involved. Emphasis should be placed on the fact that the analysis of results will focus on comparison of scores of CTN's as a group with the group scores of nurses who have not had a CTN course. It is recommended that, upon publication, the bank of objectives on which evaluations are based be made available to all nursing stations, and that the findings of the field study be forwarded to the participants. Such follow-up will allow self-evaluation and continued professional development, as well as increase the value of the evaluation in the eyes of the participants.

General Recommendations

A review of the literature has demonstrated that estimation of the reliability of measurement instruments, while necessary, constitutes an insufficient appraisal of them. It is also essential to estimate the degree to which the variable of interest is being measured--the validity of the evaluative instruments. Therefore, it is recommended that researchers involved in the health care field undertake to both estimate and report the validity of the instruments which they utilize.

A consensus regarding the minimally acceptable behavior required from members of both established and emerging professional and paraprofessional groups is necessary to:

a) validly evaluate performance, as content validity of instruments is crucial to generalizability of results; and

b) encourage rationality in the development and utilization of the proliferating types of personnel in the health care field.

Therefore it is recommended that educators and administrators attempt to delineate the competencies which are to be expected of the various levels of health care personnel, particularly new occupational groups.

The instruments and design for field evaluation which have been the foci of this study were but one phase in a sequence of projects associated with the CTN programmes. Following the identification of needs for further clinical skills, the core curriculum was rigourously built by an inter-disciplinary group. While this study has subsequently indicated that there is a change in behaviour upon completion of the course in the educational setting, it has not been shown that such change extends to the practice setting. It is recommended that field evaluation be undertaken to determine whether or not the CTN course does influence the practice behavior of nursing station nurses. Findings from such on-site assessment should certainly be available and considered when decisions regarding the continuation of the CTN courses are made.

Should use be made of any portion of the evaluation package in the future, utilizers must be cognizant of two major considerations. Firstly, the validity established is applicable only to evaluation on that curriculum delineated in the bank of objectives (cf. p. 2), and not to dissimilar curriculum content. Therefore, modifications of the instruments are necessary if skills and knowledge other than those specified in the bank of objectives are to be assessed. Secondly, due to the small sample size used during validation studies, stability of the correlation coefficients should not be assumed, although replication of this study would demonstrate whether or not such stability exists.

Within the context of the broader implications of this study, measurement validity is only one area of t[redacted]cipline of education which has relevance to the health care field. It is this investigator's opinion

(based on professional associations during and prior to this study), that educators of health personnel, while possessing content expertise, often lack solid pedagogical and measurement background. If is therefore suggested that those assuming responsibility for educational programmes develop substantial background in the broader field of education. Further, while necessarily developing theory specific to their discipline, the health professions should utilize the theories of other disciplines (of which education is but one). Such enlargement of the conceptual framework of the health professions might well enhance the health care delivered.

REFERENCES

American Psychological Association. Standards for educational and psychological tests and manuals. Washington: American Psychological Association, 1966.

Anastasi, A. Psychological testing (3rd ed.). New York: The Macmillan Co., 1968.

Bailit, H., Lewis, J., Hochheiser, L., & Bush, N. Assessing the quality of care. Nursing Outlook, 1975, 23(3), 153-159.

Bedford, M. A. Color atlas of ophthalmological diagnosis. Chicago: Year Book Medical Publishers, Inc., 1971.

Boudreau, T. J. (Chairman). Report of the committee on nurse practitioners. Ottawa: Health and Welfare, Canada, 1972.

Bracht, G. H., & Glass, G. V. The external validity of experiments. American Educational Research Journal, 1968, 5, 437-467.

Bull, T. R. Color atlas of E.N.T. diagnosis. Chicago: Year Book Medical Publishers, Inc., 1974.

Bullough, B. Is the nurse practitioner role a source of increased work satisfaction? Nursing Research, 1974, 23(1), 14-19.

Campbell, D. T., & Fiske, D. W. Convergent and discriminant validation by the multitrait-multimethod matrix. In W. A. Mehrens & R. L. Ebel (Eds.), Principles of educational and psychological measurement. Chicago: Rand McNally & Co., 1967.

Campbell, D. T., & Stanley, J. C. Experimental and quasi-experimental designs for research. Chicago: Rand McNally & Co., 1966.

Catalanello, R. F., Mingo, K. A., & Pinches, G. E. Evaluative research design for a health manpower innovation. Social Science and Medicine, 1972, 6, 229-239.

Chatton, M. J. (Ed.). Handbook of medical treatment (13th ed.). Los Altos, California: Lange Medical Pulications, 1972.

Chenoy, N. C., Spitzer, W. O., & Anderson, G. D. Nurse practitioners in primary care II. Prior attitudes of a rural population. Canadian Medical Association Journal, 1973, 108, 998-1003.

Chioni, R. M. A curriculum model for preparing tomorrow's nurse practitioners. In Challenge to nursing education, preparation of the professional nurse for future roles. New York: National League for Nursing, 1971.

Chioni, R. M., & Panicucci, C. Tomorrow's nurse practitioner. Nursing Outlook, 1970, 18(2), 32-35.

Clelland, V. Implementation of change in health care systems. Journal of Nursing Administration, 1972, 2(5), 64-69.

Conn, H. F., & Conn, R. B. Current diagnosis - 3. Philadelphia: W. B. Saunders Co., 1971.

Connelly, J. P., Stoeckle, J. D., Lepper, E. S., & Farrissey, R. M. The physician and the nurse - their interprofessional work in office and hospital ambulatory settings. The New England Journal of Medicine, 1966, 275, 765-769.

Cook, D. L. The Hawthorne effect in educational research. In W. J. Gephart & R. B. Ingle (Eds.), Educational research: selected readings. Columbus, Ohio: Charles E. Merrill Publishing Co., 1969.

Cronbach, L. J. Essentials of psychological testing (2nd ed.). New York: Harper & Row, 1960.

Cronbach, L. J. Test validation. In R. L. Thorndike (Ed.), Educational measurement (2nd ed.). Washington: American Council on Education, 1971.

Cronbach, L. J., & Meehl, P. E. Construct validity in psychological tests. In W. A. Mehrens & R. L. Ebel (Eds.), Principles of educational and psychological measurement. Chicago: Rand McNally & Co., 1967.

Cronbach, L. J., Rajaratnam, N., & Gleser, G. C. Theory of generalizability: a liberalization of reliability theory. British Journal of Statistical Psychology, 1963, 16, 137-163.

Davidson, Sir S., & McLeod, J. The principles and practice of medicine (10th ed.). Edinburgh: Churchill Livinstone, 1972.

Day, L. R., Egli, R., & Silver, H. K. Acceptance of paediatric nurse practitioners. American Journal of Diseases of Children, 1970, 119, 204-208.

deCastro, F. J., & Rolfe, U. T. An evaluation of new primary paediatric professionals. Journal of Medical Education, 1974, 49, 192-193.

Dolan, J. A. Nursing in society, a historical perspective. Philadelphia: W. B. Saunders Co., 1973.

Duncan, B., Smith, A. N., & Silver, H. K. Comparison of the physical assessment of children by paediatric nurse practitioners and paediatricians. American Journal of Public Health, 1971, 61, 1170-1176.

Ebel, R. L. Estimation of the reliability of ratings. Psychometrika, 1951, 16, 407-424.

Ebel, R. L. Obtaining and reporting evidence on content validity. In C. I. Chase & H. G. Ludlow (Eds.), Readings in educational and psychological measurement. Boston: Houghton Mifflin Co., 1966.

Ebel, R. L. Essentials of educational measurement. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1972.

Fox, D. J. Fundamentals of research in nursing (2nd ed.). New York: Appleton-Century-Crofts, 1970.

Flynn, B. C. Evaluating professional training in the human services: the case of the nurse clinician. Unpublished doctoral dissertation, University of Wisconsin, 1972.

Glass, G. V., & Stanley, J. C. Statistical methods in education and psychology. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1970.

Hazlett, C. B. Task analysis of the clinically trained nurse (C.T.N.). Nursing Clinics of North America, 1975, 10.

Hellman, L. M., & Pritchard, J. A. Williams' obstetrics (14th ed.). New York: Appleton-Century-Crofts, 1971.

Henriques, C. C., Virgadamo, V. G., & Kohane, M. D. Performance of adult health appraisal examinations utilizing nurse practitioners - physician teams and paramedical personnel. American Journal of Public Health, 1974, 64, 47-53.

Hershey, N. Standards of performance in expanded practice. In M. H. Browning & E. P. Lewis (Compilers), The expanded role of the nurse. New York: American Journal of Nursing Co., 1973.

Hospital Week. Chicago: American Hospital Association, 1973, 12(9), 1.

Houston, J. C., Joiner, C. L., & Trounce, J. R. A short textbook of medicine (3rd ed.). London: The English Universities Press Ltd., 1968.

Hughes, J. G. (Ed.). Synopsis of paediatrics (3rd ed.). Saint Louis: The C. V. Mosby Co., 1971.

Ingles, T. A new health worker. American Journal of Nursing, 1968, 68, 1059-1061.

Kempe, C. J., Silver, H. K., & O'Brien, D. Current paediatrics diagnosis and treatment (2nd ed.). Los Altos, California: Lange Medical Publications, 1972.

Kergin, D. J. (Chairman). Report of the committee on clinical training of nurses for medical services in the North. Ottawa: Health and Welfare, Canada, 1970.

Kergin, D. J., & Spitzer, W. O. An education program for nurse practitioners, report on demonstration project. Hamilton, Ontario: McMaster University, 1972.

Kergin, D. J., Yoshida, M. A., & Tidey, M. E. A study of nurse activities in primary care settings. Hamilton, Ontario: McMaster University, 1971.

Kerlinger, F. N. Foundations of behavioral research (2nd ed.). New York: Holt, Rinehart and Winston, Inc., 1973.

Krever, H. Liability for negligence. In S. R. Good & J. C. Kerr (Eds.), Contemporary issues in Canadian law for nurses. Toronto: Holt, Rinehart and Winston of Canada, Ltd., 1973.

Lambertson, E. C. Not quite MD, more than PA. Hospitals, 1971, 45(23), 70; 72; 74; 76; 136.

Lees, R. E. M. The practice nurse: a comparison of Canadian and British patient acceptance. Canadian Family Physician, 1973, 19(12), 71-77.

Letourneau, C. U. The assistant physician. Hospital Management, 1968, 105(4), 55-57.

Levene, G. M., & Calnan, C. D. Color atlas of dermatology. Chicago: Year Book Medical Publishers, Inc., 1974.

Lewis, C. E., Resnik, B. A., Schmidt, G., & Waxman, D. Activities, events and outcomes in ambulatory patient care. New England Journal of Medicine, 1969, 280, 645-649.

Linn, L. S. Expectation vs. realization in the nurse practitioner role. Nursing Outlook, 1975, 23(3), 166-171.

Lyght, C. E. (Ed.). The Merck manual of diagnosis and therapy (11th ed.). Rahway, New Jersey: Merck Sharp & Dohme Research Laboratories, 1966.

Mager, R. F. Preparing instructional objectives. Belmont, California: Fearon Publishers, 1962.

Maguire, T. O., & Hazlett, C. B. Reliability for the researcher. Alberta Journal of Educational Research, 1969, 15(2), 117-125.

McCormack, G. B. The visiting nurse becomes a nurse practitioner. Nursing Outlook, 1974, 22(2), 119-123.

Medical Services Branch. Report on the clinical training of nurses program with respect to the re-inclusion of McGill University. Ottawa: Health and Welfare, Canada, 1974.

Moore, A. C. Nurse practitioner: reflections on the role. Nursing Outlook, 1974, 22(2), 124-127.

Moore, M. F., Barber, J. H., Robinson, E. T., & Taylor, T. R. First-contact decisions in general practise. Lancet, 1973, 1, 817-819.

Mosier, C. I. A critical examination of the concepts of face validity. In W. A. Mehrens & R. L. Ebel (Eds.), Principles of educational and psychological measurement. Chicago: Rand McNally & Co., 1967.

Mussallem, H. K. The expanding role: where do we go from here? The Canadian Nurse, 1971, 67(9), 31-34.

Myles, M. F. Textbook for midwives (8th ed.). Edinburgh: Churchill Livingstone, 1971.

Oxorn, H., & Foote, W. R. Human labor and birth (2nd ed.). New York: Appleton-Century-Crofts, 1968.

Prior, J. A., & Silberstein, J. S. Physical diagnosis. The history and examination of the patient (4th ed.). Saint Louis: The C. V. Mosby Co., 1973.

Products in the service of medicine. Toronto: Connaught Medical Research Laboratories, 1972.

Reed, D. E., & Roghmann, K. J. Acceptability of an expanded nurse role to nurses and physicians. Medical Care, 1971, 9, 372-377.

Sackett, D. L., Spitzer, W. O., Gent, M., & Roberts, R. S. The Burlington randomized trial of the nurse practitioner: health outcomes of patients. Annals of Internal Medicine, 1974, 80(2), 137-142.

Selltiz, C., Jahoda, M., Deutsch, M., & Cook, S. W. Research methods in social relations (Rev. ed.). New York: Holt, Rinehart and Winston, 1959.

Sidel, V. W. Feldshers and feldsherism, the role and training of the feldsher in the USSR. The New England Journal of Medicine, 1968, 278, 934-940; 987-992.

Silver, H. K., & Duncan, B. Time-motion study of paediatric nurse practitioners; comparison with "regular" office nurses and paediatricians. The Journal of Paediatrics, 1971, 79, 331-336.

Sjoberg, G., & Nett, R. A methodology for social research. New York: Harper & Row, Publishers, Inc., 1968.

Smale, S. R. Activities of nurses and physicians in ambulatory care. Nursing Clinics of North America, 1971, 6(3), 489-497.

Spitzer, W. O., & Kergin, D. J. The nurse practitioner: calling the spade a spade. Ontario Medical Review, 1971, 38(6), 165.

Spitzer, W. O., & Kergin, D. J. Nurse practitioners in primary care I. The McMaster University educational program. Canadian Medical Association Journal, 1973, 108, 991-995.

Spitzer, W. O., Kergin, D. J., Yoshida, M. A., Russell, W. A. M., Hackett, B. C., & Goldsmith, C. H. Nurse practitioners in primary care III. The southern Ontario randomized trial. Canadian Medical Association Journal, 1973, 108, 1005-1016.

Spitzer, W. O., Sackett, D. L., Sibley, J. C., Roberts, R. S., Gent, M., Kergin, D. J., Hackett, B. C., & Olynich, A. The Burlington randomized trial of the nurse practitioner. The New England Journal of Medicine, 1974, 290, 251-256.

Stead, E. A. Jr. Conserving costly talents - providing physicians' new assistants. Journal of the American Medical Association, 1966, 198(10), 182-183.

Taber, C. W. Taber's cyclopedic medical dictionary (7th ed.). Philadelpia: F. A. Davis Co., 1956.

Talso, P. J., & Remenchik, A. P. Internal medicine based on mechanisms of disease. Saint Louis: The C. V. Mosby Co., 1968.

Thorndike, R. L., & Hagen, E. H. Measurement and evaluation in psychology and education (2nd ed.). New York: John Wiley and Sons, Inc., 1961.

Williams, C. A. Nurse practitioner research: some neglected issues. Nursing Outlook, 1975, 23(3), 172-177.

APPENDIX A

Written General and Paediatrics Examination

(Answers and scoring key for written examinations available from author or Medical Services Branch, Health and Welfare, Canada.)

I.D. No. ☐☐

WRITTEN EXAMINATION

GENERAL: ADULTS AND PAEDIATRICS

INSTRUCTIONS TO NURSES:

1. In questions where the signs or symptoms are called for, the following definitions apply:

   i) Sign - evidence of abnormality upon examination of patient.

   ii) Symptom - abnormality which the adult patient notices about himself, or which the paediatric patient's parent notices.

2. Multiple Choice questions: Circle the letter of the one best answer, unless otherwise indicated.

1. Which type of management of female contacts should be used for the control of gonorrhea?

   a) a cervical smear, VDRL, and treatment whether they are positive or negative
   b) a cervical smear, VDRL, and treatment if either is positive
   c) a cervical smear, and treatment whether it is positive or negative
   d) a cervical smear, and treatment if it is positive

2. Somatic abdominal pain (as opposed to visceral abdominal pain) is:

   a) often described as vague, aching and deep seated
   b) sharp, and well localized in one area
   c) sometimes referred to the shoulder
   d) transmitted via the autonomic nervous system

3. Which of the following is <u>not</u> associated with iron deficiency anemia?

   a) cardiac failure
   b) a heart murmur
   c) an enlarged liver
   d) an enlarged spleen

4. Three of the following are excellent sources of dietary iron. Which is <u>NOT</u>?

   a) eggs
   b) "fortified" cereals
   c) bread made with "fortified" flour
   d) leafy green vegetables

5. If a patient has a large, contaminated laceration with tendon damage and has not had tetanus immunization, what should the nurse do before evacuating him?

   a) give tetanus immune globulin, but not suture the wound
   b) give tetanus immune globulin, and simple skin closure
   c) give tetanus toxoid and simple skin closure
   d) give tetanus toxoid, but not suture the wound

6. The three kinds of deafness are:

   a) sensory, conductive, and bone
   b) sensory, conductive, and mixed
   c) congenital, neoplastic, and conductive
   d) congenital, conductive, and sensory

7. The action of tolbutamide and closely related drugs is to:

   a) stimulate the pancreas to put out more insulin
   b) increase the glucose absorbing properties of the cell
   c) inhibit the action of glucose transaminase
   d) stimulate metabolism by providing an artificial form of insulin

8. With which complaint will males with syphilis commonly present?

   a) a painless, indurated, punched out lesion located in the genital area
   b) edema, tenderness, and induration of the glans and urethra
   c) a fluid-filled scrotal sac surrounding the testis or located in the spermatic cord
   d) thick, copious, yellow urethral discharge

9. Which condition is <u>not</u> a complication of streptococcal tonsillitis?

   a) a peritonsillar abscess
   b) rheumatic fever
   c) pyelonephritis
   d) otitis media

10. An adult who has had previous episodes of acute bronchitis and presents with a runny nose, cough, but no sputum, should be treated with:

   a) tetracycline if febrile, and a broncho-dilator if wheezing
   b) tetracycline whether febrile or not, and a broncho-dilator if wheezing
   c) tetracycline if febrile, and a broncho-dilator whether wheezing or not
   d) tetracycline whether febrile or not, and a broncho-dilator whether wheezing or not

11. The obese adult diabetic

   a) often produces insulin at high levels
   b) has polyuria in reaction to his polydipsia
   c) starts to spill sugar into his urine with a blood sugar of about 120 mgm%
   d) all of the above
   e) none of the above

12. The patient with an initial acute myocardial infarction should be managed by being placed in a sitting position, given oxygen at 6 litres per minute:

   a) given a narcotic such as morphine and evacuated immediately
   b) given a narcotic such as morphine, and maintained at the nursing station unless signs of pulmonary edema or congestive heart failure develop
   c) given a mild analgesic (so as to not depress respirations), and maintained at the nursing station unless signs of pulmonary edema or congestive heart failure develop
   d) given a mild analgesic (so as to not depress respirations), and evacuated immediately

13. Patients presenting with mild symptoms of pelvic inflammatory disease throughout the menstrual cycle should be suspected of having:

   a) Ca in situ
   b) endometriosis
   c) gonorrhea
   d) syphilis

14. i) On examination, the patient with emphysema will usually have a percussion note which is:

a) normal
b) vesicular
c) hyperresonant
d) hyporesonant

ii) and his expiratory phase will be:

a) shortened
b) normal
c) prolonged

15. The presenting complaint of patients with intermittent claudication will be _____(symptom) _____(site), which they describe as being relieved by _____.

16. Chronic peripheral arterial disease is aggravated by prolonged exposure to heat or cold, trauma, and excessive use of the extremity. Patients with the disease should be advised to avoid these. In addition, list three other personal habits which should be discussed during patient education:

a) _____

b) _____

c) _____

17. i) Adult patients with chronic lung disease (especially emphysema) should be given oxygen at a maximum rate of _____ litres/minute.

ii) The sound heard on auscultation produced by fluid in the alveoli is known as a:

a) souffle
b) rhonchus
c) râle
d) bruit

iii) The sound heard on auscultation produced by narrowing of the bronchi is known as a:

a) souffle
b) rhonchus
c) râle
d) bruit

18. List four life-threatening conditions which patients develop as complications of long-standing hypertension. (Do not include aneurysm, as medical authorities do not agree on their causations.)

a) _____ c) _____

b) _____ d) _____

19. List five complications of acute otitis media other than hearing loss.

a) _____ d) _____

b) _____ e) _____

c) _____

20. Tachycardia may be present in patients on digitalis therapy. Digitalis toxicity may also be manifested by:

a) _____ c) _____

b) _____

21. i) The side effects of diuretics include lowered serum _____.

ii) Circle the TWO most commonly occurring side effects of diuretics:

a) fatigue d) elevated white blood count

b) vertigo e) muscle weakness

c) petechial rash f) retinal neuropathy

22. i) A patient with an acute myocardial infarction will usually present with:

a) blood pressure which is _____.

b) an apical beat which is _____

and/or _____.

ii) Signs of pulmonary edema may be superimposed. The outstanding signs of frank pulmonary edema are:

a) on auscultation, _____ over the long bases

b) on inspection, _____

23. The adult patient with the chronic picture of rheumatic fever may be detected by finding, on examination, _____ resulting from _____. The treatment for this is _____.

24. With any of the groups of signs and symptoms listed below, the patient's presenting complaint is abdominal pain. From the following list of diagnoses, select the one which best fits each clinical picture, and enter the letter in the blanks.

Diagnoses

a) pyelonephritis
b) urinary calculus
c) ectopic pregnancy
d) pelvic inflammatory disease
e) mesenteric adenitis
f) diverticulitis
g) pancreatitis
h) infectious hepatitis
i) appendicitis
j) bowel obstruction

_____ colicky, periumbilical pain which progresses to more constant pain, vomiting, and increased bowel sounds in the early stages

_____ severe abdominal and/or flank pain, vomiting, fever, rigors

_____ sudden pain in the epigastrium or L.U.Q. radiating to the back

_____ acute cutting pain beginning in either upper quadrant and radiating downward

_____ recurrent, spasmodic pain low on either side of the abdomen with or without pelvic cramping and shoulder pain

_____ sharp pain unilaterally or bilaterally in the lower abdomen, aggravated by defecation or movement of the cervix

_____ pain beginning in the epigastrium or pen-umbilical area, localizing to the R.L.Q.

_____ crampy abdominal pain with tenderness in the left lower quadrant

25. A person in shock from any cause may have pallor, cold and clammy skin, hypotension, and a pulse which is weak, rapid and thready. Circle those items in the list below which may occur in anaphylactic shock.

a) weakness
b) cyanosis
c) nausea
d) vomiting
e) impairment of vision
f) impairment of hearing
g) urticaria
h) dyspnea
i) giddiness
j) edema
k) bronchospasm
l) peripheral vascular collapse
m) convulsions
n) coma
o) cardiac arrest

26. Therapy for scalp ringworm (tinea capitis) should include the oral medication _____, which should be _____ _____ if there is no improvement in two weeks.

27. If a patient with otitis media still has acute symptoms after 48 hours on an antibiotic, the medication should be _____.

28. i) Specifically, which gram positive organisms usually cause otitis media?

a) _____

b) _____

c) _____

ii) If the causative organism of otitis media is gram negative, specifically which organism is it most likely to be?

_____

29. The condition in which insufficient cardiac output leads to blood pooling in the venous system is known as _____.

30. For adults with iron deficiency anemia, the usual treatment is ferrous gluconate, or ferrous _____. This medication should be given _____(route), in a dosage of _____ given _____(frequency).

31. i) When doing a Pap smear, a _____ should be placed on the _____, and rotated through 360°.

ii) However, if the endocervix is not visible, a _____ should be rotated through 360° after placing it _____ _____.

iii) A specimen should not be collected from the vaginal pool, because _____.

32. i) An average adult's basic (maintenance) fluid requirement is between _____ ml and _____ ml in 24 hours.

ii) If given intravenously, this should be in the ratio of _____ part(s) of 5% G/NS (glucose in normal saline) to _____ part(s) of 5% G/W (glucose in water) over a 24 hour period.

iii) Potassium should <u>not</u> be added to the intravenous if the patient _____.

TRUE OR FALSE: Place T or F on the line to the right of each statement.

33. Signs of right-sided cardiac failure include increased jugular venous pressure, enlargement of the liver, and peripheral edema. _____

34. A blood sugar of 40 mgm% or less is hypoglycemia in an adult or infant older than the newborn. _____

35. The solution which should be used to irrigate an ear which is, or is suspected of being perforated is normal saline (sterile). _____

36. Obese individuals have a higher percentage of their body weight in water than lean individuals. _____

37. Frequent examination of the throat (with tongue blade and light) is one of the best methods of monitoring the condition of a child with epiglottitis, or suspect epiglottitis. _____

38. List five signs (exclusive of laboratory findings) associated with urinary tract infection in the newborn.

a) _____ d) _____

b) _____ e) _____

c) _____

39. List the four main causes of major motor seizures in children.

a) _____ c) _____

b) _____ d) _____

40. The drugs of choice for urinary tract infection in children over one month of age are either _____ or _____ in the appropriate dosages.

41. The drug of choice for the treatment of streptococcal tonsillitis is _____, which should be given for _____ days in an appropriate dose.

42. Listed below are diagnoses and treatments.

Diagnosis	Treatment
1. laryngotracheobronchitis	a. aminophylline
2. epiglottitis	b. ampicillin
3. viral bronchopneumonia	c. penicillin and cloxacillin
4. staphylococcal bronchopneumonia	d. oxygen & adequate hydration
5. bronchiolitis	e. cool, moist air
6. bronchitis	f. solucortef

For each of the children described below with respiratory problems, select:

i) the corresponding number of the most likely diagnosis

ii) the corresponding letter for the treatment which should be started immediately.

	Diagnosis	Treatment
A 1 year old with an increased respiratory rate, indrawing without stridor, inspiratory rales, and an expiratory wheeze.	_____	_____
A 6 month old infant with fever of rapid onset, cough, diffuse rales, increased respiratory rate, and rapidly progressive deterioration.	_____	_____
A 2 1/2 year old quiet, toxic-looking child with rapid onset of fever, deep but slow respirations, and who is drooling.	_____	_____
A 2 1/2 year old who has had a runny nose and upper respiratory infection for two or three days. He now has a low grade fever, inspiratory stridor, indrawing, and a harsh cough.	_____	_____

43. A child has gastroenteritis and dehydration, is weak and drowsy. The nurse should look for other signs of acidosis which are:

a) an acid urine

b) _____ in the urine.

In addition, the nurse should do a physical examination, examining specifically for signs of acidosis which are:

a) _____

b) _____

44. i) What percent dehydrated is the child who is:

a) mildly dehydrated? _____ %

b) moderately dehydrated? _____ %

c) severely dehydrated? _____ %

ii) The three cardinal signs of dehydration should be checked for in a child who is suspected of being dehydrated. Below is a series of statements related to dehydration in children. Place T or F on the line to the right of <u>each</u> statement.

	T or F
a) A mildly dehydrated child will have soft and sunken eyes.	_____
A mildly dehydrated child will have dry mucous membranes.	_____
A mildly dehydrated child will have decreased blood pressure.	_____
b) A moderately dehydrated child will have soft and sunken eyes.	_____
A moderately dehydrated child will have dry mucous membranes.	_____
A moderately dehydrated child will have decreased blood pressure.	_____
c) A severely dehydrated child will have soft and sunken eyes.	_____
A severely dehydrated child will have dry mucous membranes.	_____
A severely dehydrated child will have decreased blood pressure.	_____

45. Sodium bicarbonate (7.5%) I.V. should be given to a moderately to severely dehydrated child in the dosage of:

a) 2.5 mEq/kg of body weight over the first 3 hours of treatment
b) 2.5 mgm/kg of body weight over the first 3 hours of treatment
c) 0.5 mEq/kg of body weight over the first 3 hours of treatment
d) 0.5 mgm/Kg of body weight over the first 3 hours of treatment

46. A newborn whose hips "click" on abduction when flexed has:

a) avascular necrosis of the head of the femur
b) hypertrophy of the acetabulum
c) dysplasia of the hips
d) ischemic contracture of the hips

47. The newborn whose hips "click", but will abduct when flexed:

a) will likely have no problems, but should be carefully examined when he starts to walk
b) should be examined regularly, and treatment commenced if hip abduction becomes difficult, or he develops an extra thigh fold
c) should have treatment initiated and then be flown out on the next scheduled flight for further assessment
d) should have no treatment initiated, but be flown out on the next scheduled flight for further assessment

48. A nurse should book evacuation on the next scheduled flight of the child in whom which of the following types of seizures are suspected?

a) major motor or minor motor
b) major motor or psychomotor
c) minor motor or psychomotor
d) all of the above

49. A laceration or wound large enough that it would ordinarily require suturing should <u>not</u> be sutured if:

a) _____

b) _____

c) _____

50. What should be done, in addition to consulting a physician, if a laceration is suspected of involving tendon, muscle, nerve or artery damage?

_____

51. What protection against tetanus should be given to a patient with a:

a) severe contaminated wound who has not been immunized against tetanus? _____

b) severe, contaminated would who had a tetanus series seven years ago? _____

c) minor wound who has never had tetanus immunization?

_____

d) minor wound who has kept his tetanus immunization up to date and had a booster seven months ago? _____

APPENDIX B

Written Obstetrics and the Newborn Examination

(Answers and scoring key for written examinations available from the author or Medical Services Branch, Health and Welfare, Canada.)

I.D. No. ☐☐

WRITTEN EXAMINATION:

OBSTETRICS AND THE NEWBORN

INSTRUCTIONS TO NURSES:

1. In questions where signs or symptoms are called for, the following definitions apply:

   i) Sign - evidence of abnormality upon examination of the patient

   ii) Symptom - abnormality which the patient notices about herself

2. Multiple Choice questions: Circle the letter of the one best answer, unless otherwise indicated.

1. The relationship of the long axis of the fetus to the long axis of the mother is called the:

   a) position
   b) lie
   c) presentation
   d) rotation

2. Engagement in a vertex presentation has occurred when the:

   a) widest diameter of the presenting part has passed through the brim of the pelvis
   b) vertex is at station 0
   c) vertex is at the level of the ischial spines
   d) all of the above

3. In a normal labor, L.O.A. position, following descent there should be:

   a) increasing flexion of the fetal head
   b) increasing extension of the fetal head
   c) rotation of the sagittal suture to the transverse diameter
   d) rotation of the occiput to the posterior

4. Which of the following would <u>not</u> be considered a high risk pregnancy?

   a) a previous stillbirth
   b) a previous early abortion
   c) previous phlebitis
   d) maternal diabetes
   e) none of the above

5. A patient is taking a certain strength of a particular oral contraceptive, but has breakthrough bleeding. Her medication should either be changed to another oral contraceptive with a different combination of estrogen and progesterone <u>or</u>:

   a) continued for at least another month
   b) continued, but the pattern of taking it altered to coincide with the new menstrual cycle
   c) continued for at least another month, and if the same thing happens again, discontinued and another method of contraception initiated
   d) none of the above

6. A woman has had amenorrhea since her last delivery. On examination, the nurse decides she is pregnant again, and is able to palpate the fundus midway between the symphysis pubis and umbilicus. What gestation is this pregnancy likely to be?

   a) 12 weeks

   b) 16 weeks

   c) 20 weeks

   d) 24 weeks

7. A pregnant woman had her L.N.M.P. on November 30, 1974. Her expected date of confinement is _____.

8. Four pregnant women present at prenatal clinic with the following histories of deliveries. Calculate their gravidity and parity.

   Mrs. A. - 1 set of twins at 39 weeks gestation
   - 1 livebirth at 32 weeks gestation
   - 1 stillbirth at 36 weeks gestation G_____ P_____

   Mrs. B. - 1 abortion at 12 weeks gestation
   - 1 livebirth at 28 weeks gestation, but babe died when 6 hours old
   - 1 stillbirth at 33 weeks gestation G_____ P_____

   Mrs. C. - 1 miscarriage at 12 weeks gestation
   - 1 set of twins at 40 weeks gestation
   - 2 livebirths at 40 weeks gestation G_____ P_____

   Mrs. D. - 1 livebirth at 29 weeks gestation
   - 1 livebirth at 39 weeks gestation
   - 1 miscarriage at 12 weeks gestation G_____ P_____

9. Circle the names of the women described in question 8. who should be considered at risk for this pregnancy.

   Mrs. A. Mrs. C.

   Mrs. B. Mrs. D.

10. Should an obstetrical patient with glucose present in her initial urine specimen be referred to a physician? _____

    Common causes of glucose in the urine of a pregnant woman are:

    1) _____

    2) _____

    3) _____

11. If a patient is 5 ft. 0 in. or less in height, she is more likely than taller women to have _____, which will lead to a difficult and complicated delivery.

    The risks are less, however, if the woman's obstetrical history reveals

    _____.

12. The signs of pre-eclampsia include an elevated blood pressure greater than _____/_____, or a rise of _____ mm systolic or _____ mm diastolic greater than the patient's normal blood pressure.

13. A pregnant woman is considered to be anemic if her hemoglobin is _____ or less.

14. If rapid weight gain, proteinuria and an elevated blood pressure are considered signs of pre-eclampsia, what symptoms of pre-eclampsia should patients be asked to report immediately?

    1) _____ 4) _____

    2) _____ 5) _____

    3) _____

15. At prenatal clinic two weeks ago you saw four multiparous women. Each was at 8 weeks gestation. Last night each of them came to the nursing station with "bleeding". Additional information is given below for each. Give the most likely diagnosis for each.

	DIAGNOSIS
Mrs. A. has slight dark brown discharge, severe abdominal pain, an acute abdomen, and is going into shock.	_____
Mrs. B. has slight bleeding which started yesterday morning, and mild abdominal cramps.	_____
Mrs. C., following intercourse, had bleeding which has since stopped. She has had no abdominal pain, but has increased vaginal discharge.	_____
Mrs. D. has severe bleeding, has passed some tissue, and continues to pass clots. She has a dilated cervix, and severe abdominal pain.	_____

16. In a vertex delivery, firm manual control of the fetal head at the vulva is necessary to prevent:

    1) _____

    2) _____

17. a) The indication that the fetal shoulders are positioned for delivery is _____.

    b) The _____ shoulder should be delivered first.

18. a) The _____ and _____ must be examined after their delivery as retained portions can predispose the mother to _____ or _____.

    b) The number of vessels normal in an umbilical cord is _____ vein(s) and _____ artery(ies).

19. An episiotomy should be performed _____ contraction(s), after allowing the patient to push for several contractions when the presenting part is extending the vulva.

    The indications for an episiotomy are:

    1) forcep delivery 4) _____

    2) _____ 5) _____

    3) _____ 6) _____

20. List the signs of placental separation which occur before the placenta appears in the vagina in an uncomplicated delivery.

    1) _____

    2) _____

    3) _____

21. What are the causes of failure of the uterus to contract following delivery of the babe and placenta?

    1) _____ 3) _____

    2) _____ 4) _____

22. a) The two major postnatal causes of peripheral cyanosis of the newborn without central cyanosis are:

    1) _____

    2) _____

b) Differentially from peripheral cyanosis, central cyanosis may be readily noted even with the baby warmly bundled, as the _____ area and _____ will be cyanotic.

c) Consider that one major condition leading to central cyanosis due to inadequate oxygen supply is respiratory distress (atelectasis being a specific example of it). List four other major conditions of the newborn associated with cyanosis.

1) _____ 3) _____

2) _____ 4) _____

23. Caput succedenum and cephalhematoma are two distinctly different conditions.

1) Which letter from the list below best describes caput succedenum? _____

Which letter from the list below best describes cephalhematoma? _____

a) accumulation of blood beneath the periostium

b) effusion of fluid beneath the periostium

c) accumulation of blood interstitially in the scalp

d) effusion of fluid interstitially in the scalp

e) accumulation of blood beneath the dura mater

f) effusion of fluid beneath the dura mater

2) Complete column B by placing the letter from column A which best describes the usual chronology.

A. Time Following Birth		B
a) at birth	Cephalhematoma first noticeable	_____
b) 3-4 hours	Cephalhematoma resorbs	_____
c) 5 days	Caput succedenum first noticeable	_____
d) 10 days	Caput succedenum resorbs	_____
e) 3 weeks		
f) 4 weeks or longer		

3) On examining the infant's head, there are two findings which differentiate caput succedenum from cephalhematoma. Describe these, stating how they distinguish between caput and cephalhematoma.

a) _____

_____

b) _____

_____

4) If a cephalhematoma is present, the babe should have laboratory work done to watch for a low _____ and/or a high _____ which can result.

24. a) Jaundice appearing less than 24 hours after birth is probably due to _____ or _____.

b) Later jaundice of the newborn may rarely be due to abnormal red cell metabolism, extravascular hemorrhage, or severe liver or bile duct defects.

List four more common causes of jaundice of the newborn appearing more than 24 hours after birth.

1) _____

2) _____

3) _____

4) _____

25. Fill in the Apgar scoring chart below. List the five parts in the first column and indicate the criteria used for scoring in the other three.

APGAR	0	1	2

APPENDIX C

Physical Examination Instrument

ROUTINE COMPLETE PHYSICAL EXAMINATION

I.D. Number ☐☐

GENERAL

weight	☐		☐	pulse rate	☐	≥ 1 min.	☐
height	☐		☐	pulse rhythm	☐		☐
blood pressure	☐	sitting or lying and standing	☐	respiratory rate	☐		☐
temperature	☐	oral ≥ 2 min.	☐	hemoglobin	☐	Spencer	☐
				urinalysis	☐		☐

HEAD AND NECK

General

palpates: skull	☐	generally	☐
mastoid processes	☐		☐
inspects: hair	☐		☐
scalp	☐		☐

Eyes

inspects lids, conjunctiva, pupils, etc.	☐	lids retracted	☐
examines discs	☐	ophthalmoscope held in right hand for right eye and the reverse	☐
tests: visual acuity	☐	Snellen chart at 20', one eye covered, with and without glasses	☐
pupillary light reaction	☐	room darkened, light brought in from side	☐
muscle balance	☐	6 cardinal positions	☐
visual fields	☐	confrontation test	☐
for strabismus	☐	either pupillary light reflection or cover test	☐

Nose

examines nose	☐	external nares, anterior aspects, and for patency of nostrils	☐
palpates nostrils	☐		☐

Ears

inspects pinna	☐	all aspects	☐
examines auditory canal, drums	☐	pulling auricle upward and backward, uses otoscope	☐
tests functional hearing	☐	patient occludes one ear with finger, nurse positions herself 20-30 feet away and covers her lips. Softly speaks two syllable words.	☐

Mouth and Throat

inspects: teeth, gums	☐	must separate cheeks and tongue from gingival surface, removes dentures	☐
tongue, lips	☐	using tongue blade & light	☐
pharynx, tonsils, palate	☐		☐

Neck

tests mobility	☐	asks patient to place chin on chest, turn head to each side	☐
inspects anterior neck	☐	head in moderate extension	☐
palpates neck	☐	posteriorly, laterally, over trachea and thyroid, carotid arteries	☐

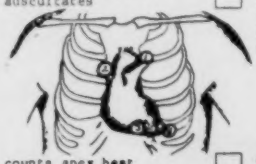
CHEST: RESPIRATORY

inspects respiratory movement	☐	anterior, posterior, lateral walls	☐
palpates: thoracic expansion	☐	hands spread with thumbs together medially and fingers over lateral wall. During both quiet and deep inspiration.	☐
fremitus	☐	with palmar surface of fingers or ulnar aspect of hand, compares corresponding areas of sides of chest, both front and back, while patient speaks, e.g., "99"	☐

percusses: anterior chest	☐	by either mediate or immediate method with finger parallel to rib interspaces. Compares one side to the other.	☐
posterior chest	☐		☐
lateral chest	☐		☐
auscultates: back	☐	compares corresponding areas of both sides while patient breaths with mouth open.	☐
anterior chest	☐		☐
lateral chest, axillae	☐		☐

CHEST: CARDIOVASCULAR

inspects precordium	☐	tangentially views for pulsations, etc.	☐
palpates for apex beat, thrills, pulsations	☐	over the precordium	☐
auscultates	☐	1) pulmonic area	☐
		2) aortic area	☐
		3) tricuspid area	☐
		4) mitral	☐



counts apex beat	☐	≥ 20 seconds	☐
measures apex position	☐	distance from midsternum	☐
palpates pulses	☐	brachial, radial	☐
		femoral, popliteal, posterior tibial, dorsalis pedis	☐

ABDOMEN

Performs inspection, auscultation, percussion and palpation in that order			☐
inspects	☐	patient supine, abdomen bared, having voided prior to exam.	☐
auscultates	☐	all quadrants, stethoscope diaphragm lightly against abdomen	☐
percusses: abdomen	☐	all quadrants	☐
liver	☐	upper border	☐
palpates gently	☐	over entire abdomen	☐
palpates deeply: for liver	☐	working up from below, asking patient to take a deep breath	☐
for spleen	☐	(L) hand over costo-vertebral angle exerting pressure, (R) hand under (L) anterior costal margin. Patient takes deep breath.	☐
both flanks	☐		☐
subrapubic region	☐		☐
LLQ, RLQ, RUQ, LUQ medially	☐		☐
groins	☐		☐
for inguinal hernia	☐	patient standing, places finger in inguinal ring, instructs him to cough	☐
performs rectal exam	☐	includes inspection of perianal area	☐

BREASTS

inspects	☐	patient sitting: arms at side, overhead, and hands pressed on hips.	☐
palpates	☐	all aspects with patient supine, and sitting with arms by side as well as arms overhead.	☐
palpates for nodes	☐	supra and infraclavicular areas, axillae	☐

MALE GENITALIA

inspects: penis	☐	foreskin retracted,	☐
scrotum	☐	patient both supine and standing	☐
palpates penis, testes	☐		☐

FEMALE GENITALIA

external genitalia	☐	inspects perineum, labia spread	☐
		palpates Bartholin's glands	☐
tests pelvic support	☐	palpates posterior and lateral vaginal walls for resistance	☐
takes Pap smear	☐	using speculum without lubricant, rotates spatula on cervical os, spreads on slide and fixes	☐
takes cervical culture	☐	with cotton-tipped swab on cervical os	☐
examines cervix, vagina	☐	before and during withdrawal of speculum	☐

MUSCULOSKELETAL

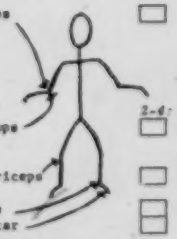
Upper extremities			
inspects: entire upper extremities	☐	all aspects must be visible to nurse	☐
for tremor	☐	with hands and arms outstretched	☐
palpates: hands, wrists	☐	through full range of motion	☐
elbows	☐		☐
shoulders	☐		☐
Lower extremities			
inspects entire lower extremities	☐	with patient standing, all aspects visible	☐
palpates: feet, ankles	☐	through full range of motion	☐
knees	☐		☐
hips	☐		☐
Spinal column			
inspects spinal column	☐	patient standing and bending from waist	☐
palpates spinal column	☐	patient bending from waist	☐
inspects posture, gait	☐	instructs patient to walk across room and back	☐

NEUROLOGICAL

Cranial Nerves

I	Olfactory	☐	asks patient to identify pungent smells such as tobacco, soap ☐
V	Trigeminal	☐	palpates masseter muscles while patient clenches teeth ☐
			tests superficial sensation over cheeks with sharp/dull discrimination ☐
			tests corneal reflex with cotton wool ☐
VII	Facial	☐	asks patient to show teeth, purse lips ☐
			asks patient to look upward and wrinkle forehead ☐
VIII	Acoustic	☐	tests bone conduction: tuning fork held on mastoid process until patient no longer hears it, then held next to ear ☐
IX	Glossopharyngeal	☐	tests gag reflex: touches tongue blade to posterior wall of pharynx or posterior third of tongue ☐
X	Vagus	☐	observes soft palate for equal bilateral movement while patient says "Ah" ☐
XI	Spinal accessory	☐	asks patient to rotate head against resistance, elevate shoulders against resistance ☐
XII	Hypoglossal	☐	asks patient to stick out tongue and wag it from side to side ☐

Reflexes



			compares one side to the other ☐
1.	biceps	☐	instructs patient to contract muscle slightly. Places one finger on tendon, and taps it with reflex hammer. Supports arm. ☐
		2-6: taps directly over tendon	
2.	triceps	☐	supports upper arm, forearm swinging freely, or folds arm across chest ☐
3.	quadriceps	☐	with leg dangling, strikes just below patella ☐
4.	ankle	☐	with knee flexed ☐
5.	plantar	☐	slowly and firmly strokes blunt instrument up lateral aspect of sole, and then across to big toe ☐

Muscle tone

tests extremities	☐	passively moves limbs through full range of motion ☐	

Muscle power

upper extremity	☐	compares bilateral hand grips ☐	
		attempts to flex or extend wrists while patient resists ☐	
		attempts to straighten arm while patient holds forearm at right angle to upper arm ☐	
lower extremity	☐	pushes down on foot while patient holds it in dorsi-flexion ☐	
		presses down on patient's straight-raised leg ☐	
trunk	☐	asks patient to sit up from supine position ☐	

For items below, instructs patient to:

Balance: tests	☐	stand with feet together, arms at sides, and eyes closed ☐	

Muscle coordination

upper extremity	☐	perform rapid rhythmic movements, such as: patting knee with hand, alternately using palm and back of hand ☐	
		move index finger back and forth between his nose and nurse's index finger ☐	
lower extremity	☐	wiggle toes rapidly up and down ☐	
		place heel on knee and run it quickly down tibial surface in straight line ☐	

Sensation

tests: sharp/dull	☐	identify sensations as nurse moves object up anterior aspects and posterior aspects of limbs and trunk, and over face ☐	
touch	☐	detect light touches with piece of cotton wool over above-mentioned surfaces ☐	
position sense	☐	detect whether toes or fingers being moved passively 'up' or 'down', nurse holding the lateral aspect ☐	

APPENDIX D

Adult History Instrument

COMPLETE HISTORY: ADULT

E: information elicited
R: information recorded

I.D. No. ☐☐

PATIENT PROFILE

E R

*name
*sex
*race
*place of birth
*date of birth
*address/band number
home facilities
time spent in present location
travel in past out of country
*religion
*marital status
occupation
*school/education
financial resources
habits: drugs
tobacco
alcohol
other
average 'day in the life of'
*name of next of kin
occupation of next of kin

CHIEF COMPLAINT

presenting symptoms and their nature
location
chronology
duration
aggravating factors
alleviating factors
associated features
degree of disability related to complaint
patient's and/or family's interpretation of symptoms in light of their past experience with similar symptoms

If this is a previously established illness:

*basis for diagnosis
medication prescribed:
*name
*dose
frequency taken
*other treatment prescribed
specific inquiry regarding appropriate effects of medication or treatment
ensures that patient understands medical terms by outlining the signs and symptoms when lack of understanding is apparent or could be a problem

PAST HISTORY

E R

*mumps
*measles
*chicken pox
*scarlet fever
*rheumatic fever
*rubella
*T.B.
other major illnesses
injuries
operations

E R

hospitalizations
allergies
transfusions
previous medical care
ongoing medications
name
dose
frequency
*immunization status
date of last chest x-ray

If patient has had a major illness, operation, or injury, elicits information regarding:
date and diagnosis
current health status in relation to the episode

FAMILY HISTORY

	E R	E R	E R	E R	E R
*year of birth present health or *date and cause of death	Mother	Father	Siblings	Spouse	Children

E R

*consanguinity
diabetes
asthma
allergy
eczema
hypertension

E R

T.B.
mental retardation
convulsions
mental disorder
other major health problems with hereditary transmission

If familial predisposition to disease indicated, appropriately elicits more detailed information

Endeavors to elicit cues regarding the patient's feelings regarding his family status

RECORD KEEPING

Attempts to obtain summaries of previous medical care
Lists and dates problems at the end of notes
Checks with patient regarding the activity of problems previously noted
Marks problems as inactive appropriately
Ensures that successful management of problems has been noted
Outlines plans for management of each active problem

FUNCTIONAL INQUIRY--whether have had any problems with or changes in:

Constitutional (E R)
- mood
- sleep
- weakness
- fatigue
- fever
- appetite/diet
- weight
- night sweats

Hematopoetic
- bleeding
- bruising
- petechiae

Integument
- color of skin
- texture of skin
- lesions
- itching
- rash,
- hair
- nails

Head and Neck

eyes
- vision
- blurring
- diplopia
- blindspots
- photophobia
- pain
- discharge
- strabismus
- glasses

ears
- hearing
- pain
- ringing
- discharge

nose
- smell
- bleeding
- discharge
- sneezing
- itching
- stuffiness

mouth and throat
- teeth
- gums
- sore throat
- dysphagia
- voice
- taste
- tongue
- postnasal drip

general
- headache
- dizziness
- trauma
- pain in neck
- lumps in neck

Breasts
- pain
- lumps
- discharge

Respiratory (E R)
- cough
- sputum
- chest pain
- hemoptysis
- wheeze
- orthopnea
- dyspnea

Cardiovascular
- pain
- palpitations
- exertional dyspnea
- nocturnal dyspnea
- blackouts
- high blood pressure
- edema
- varicosities
- intermittent claudication

Gastrointestinal
- pain
- nausea
- vomiting
- hematemesis
- indigestion
- jaundice
- bowel habit change
- diarrhoea
- constipation
- blood in stool
- hernia
- hemorrhoids

Urinary Tract
- stream
- frequency
- nocturia
- urgency
- dysuria
- back pain
- hematuria
- oliguria
- incontinence

Female Genital System
- gravida
- para
- menses: age of onset
  - LMP
  - frequency
  - duration
  - amount of flow
  - dysmenorrhea
- postmenopausal bleeding
- vaginal discharge
- itch
- sores
- dyspaerunia
- last Pap smear
- contraception
- libido
- fertility
- sexual contacts

Male Genital System (E R)
- sores
- discharge
- testicular: pain
  - swelling
  - mass
- libido
- erection
- fertility
- sexual contacts

Musculoskeletal

back and extremities:
- trauma
- pain
- stiffness
- swelling
- joint redness
- muscle power

Central Nervous System
- coordination
- sensation
- tingling
- convulsions
- loss of consciousness
- intellectual function

Endocrine
- growth rate
- polydipsia
- polyuria
- sweating
- goitre
- salt intake
- striae
- temperature tolerance

Psychological
- hyperventilation
- absenteeism
- nervousness
- nightmares
- fears
- anxieties
- voices
- feelings of persecution

For any abnormal findings in functional inquiry, elicits further information regarding their:
- nature
- chronology
- duration
- associated features
- aggravating factors
- alleviating factors
- related disability

APPENDIX E

Paediatric History Instrument

COMPLETE HISTORY: PAEDIATRIC

E: Information elicited
R: Information recorded

I.D. Number ☐☐

PATIENT PROFILE

	E	R
*name		
*sex		
*race		
*place of birth		
*date of birth		
*address/band number		
home facilities		
time spent in present location		
travel in past out of country		
*religion		
*marital status of parents		
*school/education		
financial resources		
habits: drugs		
tobacco		
alcohol		
other		
average 'day in the life of'		
*name of next of kin		
occupation of next of kin		

CHIEF COMPLAINT

presenting symptoms and their nature ☐
location ☐
chronology ☐
duration ☐
aggravating factors ☐
alleviating factors ☐
associated features ☐
patient's and/or family's interpretation of symptoms in light of their past experience with similar symptoms ☐
degree of disability related to complaint ☐

If this is a previously established illness:

*basis for diagnosis ☐
medication prescribed:
*name ☐
*dose ☐
frequency taken ☐
*other treatment prescribed ☐
specific inquiry regarding appropriate side effects of medication or treatment ☐

HISTORY OF BIRTH AND EARLY CHILDHOOD

Pregnancy

	E	R
duration		
complicated by illness/-infection: when		
drugs taken: what & when		
bleeding: when		
X-rays: when		
no. of previous pregnancies		

Labour and Delivery

	E	R
duration of labour		
anesthesia used		
instrumentation/normal		

Neonatal Period

	E	R
birthweight		
general condition		
when first cried		
respiratory problems		
jaundice: onset, severity		
deformities		
convulsions		

Feeding Patterns (if < 2 yrs.)

	E	R
breast and/or formula		
cereals:		
age started		
amount		
other solids:		
when		
what		
vitamin supplement		
rate of weight gain		
significant vomiting or diarrhea		

Growth and Development

	E	R
age at which: smiled		
rolled over		
sat		
stood		
spoke		

PAST HISTORY

*mumps		allergies
*measles		transfusions
*chicken pox		previous medical care
*scarlet fever		ongoing medications:
*rheumatic fever		name
*rubella		dose
*T.B.		frequency
other major illnesses		*immunization status
injuries		date of last chest x-ray
operations		
hospitalizations		

Ensures that patient or parent understands medical terms by outlining the signs and symptoms when lack of understanding is apparent or could be a problem. ☐

If patient has previously had a major illness, operation, or injury, elicits information regarding: date and diagnosis ☐
current health status in relation to the episode

FAMILY HISTORY

	E	R	E	R	E	R
*year of birth present health OR *date and cause of death						
	mother		father		siblings	

*consanguinity		mental retardation
asthma		convulsions
allergy		mental disorder
eczema		other major health problems with hereditary transmission
hypertension		
T.B.		
diabetes		

If familial predisposition to disease indicated, appropriately elicits more detailed information. ☐

Endeavours to elicit cues regarding the patient's or parents' feelings regarding his family status. ☐

FUNCTIONAL INQUIRY--whether have had any problems with or changes in:

Constitutional (E R)
- mood
- sleep
- weakness
- fatigue
- fever
- appetite/diet
- weight
- night sweats

Integument
- color of skin
- texture of skin
- lesions
- itching
- rash
- hair
- nails

Head and Neck

eyes: vision
- blurring
- diplopia
- blindspots
- photophobia
- pain
- discharge
- strabismus
- glasses

ears: hearing
- pain
- ringing
- discharge

nose: smell
- bleeding
- discharge
- sneezing
- itching
- stuffiness

mouth and throat
- teeth
- gums
- sore throat
- dysphagia
- voice
- taste
- tongue
- postnasal drip

general: headache
- dizziness
- trauma
- pain in neck
- lumps in neck

Breasts
- pain
- lumps
- discharge

Respiratory
- cough
- sputum
- chest pain
- hemoptysis
- wheeze
- orthopnea
- dyspnea

Cardiovascular (E R)
- pain
- palpitations
- exertional dyspnea
- nocturnal dyspnea
- blackouts
- high blood pressure
- edema
- cyanosis

Hematopoetic
- bleeding
- bruising
- petechiae

Gastrointestinal
- pain
- nausea
- vomiting
- hematemesis
- indigestion
- jaundice
- bowel habit change
- diarrhoea
- constipation
- blood in stool
- hernia
- hemorrhoids

Urinary Tract
- stream
- frequency
- nocturia
- enuresis
- urgency
- dysuria
- back pain
- hematuria
- oliguria
- incontinence

Male Genital System
- sores
- discharge
- testicular: pain
  - swelling
  - mass

Female Genital System

menses: age of onset
- LMP
- frequency
- duration
- amount of flow
- dysmenorrhea
- vaginal discharge
- itch
- sores

Musculoskeletal

back and extremities: (E R)
- trauma
- pain
- stiffness
- swelling
- joint redness
- muscle power

Central Nervous System
- coordination
- sensation
- tingling
- convulsions
- loss of consciousness
- intellectual function

Endocrine
- growth rate
- polydipsia
- polyuria
- sweating
- goitre
- salt intake
- striae
- temperature tolerance

Psychological
- hyperventilation
- absenteeism
- nervousness
- nightmares
- fears
- anxieties
- voices
- feelings of persecution

For any abnormal findings in functional inquiry, elicits further information regarding their:
- nature
- chronology
- duration
- associated features
- aggravating factors
- alleviating factors
- related disability

RECORD KEEPING
- Attempts to obtain summaries of previous medical care
- Lists and dates problems at the end of notes
- Checks with patient regarding the activity of problems previously noted
- Marks problems as inactive appropriately
- Ensures that successful management of problems has been noted
- Outlines plans for management of each active problem

APPENDIX F

Suturing Instrument

I.D. No. ☐☐

SUTURING

Of a straight, clean laceration with Interrupted Sutures

☐ Disinfects the area with skin disinfectant.

☐ Anaesthesizes the area:
- with local anaesthetic not containing epinephrine
- to a distance of 1 - 2 cm. on all sides of wound
- inserts needle under skin edges (not percutaneously)
- uses 2 or 5 cc. syringe with intradermal or subcutaneous needle (25 or 27 guage).

☐ Secures haemostasis by clamping bleeding points
- with mosquito forceps or haemostats.

☐ Ligates them with ligatures of
- plain catgut
- 000 or 0000
- cut next to knot.

☐ Sponges the area dry
- with mopping rather than wiping motion.

☐ Grasps with needle holder
- a curved cutting needle
- with 000 or 0000 } or equivalent non-absorbable external suture material.
- silk sutures.

☐ Starts at centre of wound.

☐ Uses small toothed or non-toothed tissue forceps to grasp skin.

☐ Inserts needle through full thickness of skin
- from outside to inside
- 3 - 6 mm. from skin edge,
- directed to the opposite side of wound.

☐ Inserts needle through similar point on other side
- directly opposite
- from inside to outside.

☐ Ties suture
- with a one and half square knots (by either accepted one or two handed method or using instrument tie)

☐ achieving approximation of skin edges
- without inversion
- with only slight eversion
- without wrinkling of skin.

☐ Cuts end of suture to 3 - 6 mm. length.

☐ Completes closure of incision with similar sutures
- 7 - 15 mm. apart
- by successively halving remaining distances. (Each half must lie within middle third of remaining distance.)

☐ Maintains asepsis throughout procedure,
- wears gloves throughout procedure.

☐ Applies appropriate sterile dressing or collodion.

☐ Gives appropriate tetanus protection.

APPENDIX G

Instruction Manual

INSTRUCTION MANUAL

General

Prior to actually evaluating any nurses, it is important to become thoroughly familiar with the evaluation package. The improper or incorrect use of the forms will affect the reliability and validity of the evaluation.

Each instrument designed for direct observation of performance (i.e., the suturing, history and physical, and prenatal forms) should actually be used at least five times in practice sessions. During validation studies, markers found that scoring was initially difficult because they either were unaware that the items were on the sheet, or unable to locate the items quickly enough. Item location is a problem when the nurse's performance is disorganized and does not follow the logical sequence in which the marking forms are set out. Conversely, the problem of locating the items rapidly is also heightened when a well-prepared nurse progresses quickly through a number of techniques, procedures, or questions.

It is suggested that practice sessions be accomplished through:

a) use of videotapes of histories and physicals available from the University of Alberta,
b) attendance at an out-patient clinic or in a physician's practice where complete histories and physicals are done, and/or
c) visiting a nursing station which is NOT to be used in the on-site evaluation per se.

The following instruments are included in the evaluation package, those marked with an asterisk (*) having undergone validation studies.

Written	* General and Paediatrics
	* Obstetrics and the Newborn
Observation	* Adult History
	* Paediatric History
	* Complete Routine Physical Examination
	* Suturing
	Management of an Obstetrical Patient on Initial Visit
Record Review	Histories
	Physical Examination
	Management of Prenatal Care
	Otitis Media (Acute and Chronic)
Simulation	Physical Findings

When evaluating by observation, particularly, it is crucial that the nursing station environment be as near normal as possible. Thus, while interacting appropriately with the subject and her patient, the evaluator must be as uninterruptive and unobtrusive as possible.

It is conceivable that circumstances could arise which would prohibit completion of a subject's total evaluation. Should such a situation occur, a note is to be entered on the personological data form clearly indicating the reason(s) for the lack of completion. During the on-site visit, the evaluation should consist minimally of:

a) complete assessment of one female adult patient;
b) complete assessment of one male adult patient;
c) history of one paediatric patient (less than 16 years);
d) evaluation of suturing of one laceration;
e) assessment of a prenatal patient on her first visit;
f) both written examinations;
g) review of five adult, five paediatric, and five obstetrical charts;
h) review of five acute and five chronic otitis media patients' charts.

Standardization (Consistency)

In order to obtain results from which inferences can be legitimately drawn, it is essential that the maximal degree of standardization be achieved across nurses. This is difficult to accomplish during the field evaluation because of the: variety of patients seen and the disease entities with which they present; evaluator's personal interaction with each nurse; changes in the evaluator over the length of the study (maturation); and differing workloads in the various nursing stations. Following are several measures which should be taken to increase standardization.

a) The highly objective evaluation instruments must be correctly utilized. In all tests, nurses are marked right or wrong on each item. That is, each question is specific enough that the decision the evaluator makes is "Did she or didn't she?" No part marks can be awarded.
b) The evaluator must be consistent in the use of the evaluation instruments across the total group of subjects sampled. The bank of objectives stipulates the knowledge and skills which CTN graduates should have. However, it is not known how many of these objectives are met in the field performance of nurses with and without the course. To ascertain whether or not there is a significant difference in performance, therefore, the RN's must be systematically evaluated on virtually the same measures as the CTN's.
c) To reduce the effect of changes in the evaluator over the length of the project, the order in which CTN's and RN's are evaluated should be randomized.
d) Patients to be assessed should possess the personal characteristics specified.
e) The standardizing oral questions which are included for use with several of the instruments must be asked as indicated. These are designed to allow a subject to earn credit for those items for which she could otherwise not be scored because of the particular patients seen.
f) The order in which the instruments are used should be systematic and standard for all subjects (cf. p. 123).

g) While the minimal content of the evaluation is specified (cf. p. 2), as many scorings of the instruments as practicable should be obtained for each subject. Such repeated measures of the observation tests and record review (using a new copy of the form on each occasion) will lead to a better estimation of the true scores for both groups (RN's and CTN's).

The On-Site Visit

Each subject must fill in the personological data form, and have an identification number (I.D.) assigned at the beginning of the on-site visit. However, in order to ensure anonymity, no record should be kept from which it is possible to associate the identity of the subjects with their identification numbers.

During the first day of the visit, the majority of time should be spent doing record review, to accustom the nurse to your being in the nursing station. If at all possible, work in close proximity to the clinic area so that:

a) the nurse becomes comfortable with you nearby while she is seeing patients, thus reducing observer effect;
b) the community is aware of your presence as a non-participating observer; and
c) you are aware of patient flow.

Since some content of the written and simulation examinations is in common with the observation examinations, the written and simulation assessments should be done on the last two days of the visit. Unless this standardized ordering of evaluation is followed, the content of the simulation and written examinations could influence performance on the observation examinations by suggesting appropriate patient management. For example, questions on the general written exam regarding tetanus immunization would probably remind the subjects that such protection should be included in care of the patient with a laceration, thus influencing scores on any suturing which followed the written examination. In such a case, a multiple treatmeant effect could be observed (cf. pp. 29-30), and the performance rating thus not reflect the subjects' usual practice.

At the outset of the visit, present the nurse with her instructions, and ask her to inform you when an appropriate patient for a standardized evaluation arrives. Should it become apparent that the nurse is avoiding the assessment by not doing so, more direction will be required. If, during the first two days of the visit, the nurse does not do any complete histories and/or physicals (especially if the record review has indicated that she is not in the habit of doing complete patient assessments), ask her to arrange to have a patient come in for a complete history and physical. Preferrably this assessment should be done for a patient of the nurse's choice, as she is more familiar with both the patient's likely response and her own time constraints.

All subjects should be uninformed, prior to their evaluation, of the degree of thoroughness and specific content of the field evaluation. To prevent dispersal of information regarding the instruments from those evaluated to prospective testees, neither the forms nor the keyed correct responses on the written examinations should be shown to the subjects. Additionally, the subjects should be requested not to discuss the content of the written and simulation examinations with colleagues in other nursing stations who have not been evaluated. Such measures will help to prevent sensitization of the subjects, and thus improve the validity of the results.

No nurse should be informed of her scores. Rather, it must be rigorously explained that percentages are not calculated, and raw scores for any one individual are not the object of this study. However, as studies have shown that praise tends to produce improved performance and reduce anxiety, it may be helpful to comment on the favorable aspects of the subject's capabilities.

Written Examinations

All questions have been developed from the material in the bank of objectives, in which the answers are also generally found. As the bank is not considered to be all-inclusive of medical knowledge, there are some additional correct answers for the open-ended questions. Such correct responses (which have been documented in the literature) are also indicated in the keys provided for both the GENERAL AND PAEDIATRICS and OBSTETRICS AND THE NEWBORN examinations.

During validation studies it was noted that nurses who have not had the CTN course are inclined to give responses which "get at" (approximate) the answer. Such responses are not considered correct. It is essential that the same standards be applied in marking the examinations of nurses with and without the course. Throughout, each correct response is allotted one point. No partial marks may be awarded--the answer is either correct or it is not.

Although no time limit has been set for the writing of the exams, they tend to be self-limiting in the time taken, as they involve no essay types of questions. Ideally, the subject should be uninterrupted in her writing of them, allocating approximately one-and-one-half hours for each.

Complete Patient Assessment

General Instructions

1. The patients selected for complete assessment should be freely ambulatory, and preferably conversant in English. If the latter is not the case, an interpreter must be present.

2. The nurse should select the patients herself (cf. Information for Nurses and General Instructions). Encourage the nurse to handle the introduction of the situation herself: the patient's familiarity with her will be reassuring to him. Such will also prevent erosion of the authority and responsibility of the nurse.

3. The three forms used in observer evaluation of a complete patient assessment, the COMPLETE ADULT HISTORY, the COMPLETE ROUTINE PHYSICAL EXAMINATION, and RECORDING OF PHYSICAL EXAMINATION must be stapled together upon completion of the patient's evaluation.

Complete Adult History

The COMPLETE ADULT HISTORY form, arranged in a logical sequence, is to be used both for evaluation by observation, and review of the record following patient assessment. The items in the first column, marked (E) are to be checked (☑) during the interview for each piece of information elicited by the nurse. The second column, marked (R) is applicable to the recording of items, and is to be scored when reviewing the nurse's charting.

The starred (*) items may be credited for both observation and record review if the record contains that specific piece of information. No other items can be credited unless the information is elicited or volunteered at the time of the interview being evaluated, as:

a) the situations may have changed, and

b) completeness of previous recordings cannot be assumed.

The hatched (▨) items are applicable to all patients, but indicate high risk to prenatal patients (cf. Management of Prenatal Care).

Observer Evaluation: Specific Instructions

Patient profile.

- Age of the patient is insufficient for crediting date of birth. The date of birth, to the best of the patient's knowledge, must be elicited.

- Sex of the patient may be credited if recorded either previously or at the time of the assessment.

- Credit is given for race if a disc number (for Inuit) or treaty number (for Indians) is recorded or obtained.

- Home facilities are credited if the nurse makes any enquiry into living conditions, number of people in the home, etc.

Chief complaint.

-This is the most subjective area on the form, but must remain so to be generalizable to the variety of patients who will be assessed. Due to this variability in patients, it is not feasible to specify criteria for complete enquiry into any one item in the section. Credit is therefore given if in any way the subject delves into the area covered by the item. For example, a patient may complain of burning on urination: if the nurse asks about one associated feature such as frequency, she must be credited with the item 'associated features,' although you may feel she has not fully inquired into the item.

-Many of the items in the chief complaint are directly applicable to the functional inquiry. When scoring on the chief complaint, therefore, the appropriate items in the functional inquiry must be cross-referred to and scored.

Past history.

- The individual communicable diseases must be enquired about, credit not being given for a general question such as: "What childhood or communicable diseases have you had?" The reasons for this being:
  a) the patient's response will be more accurate if his memory is stimulated by specific enquiries;
  b) many members of the general public are not knowledgeable regarding which diseases are included in such a general classification; and
  c) unless specific reference is made to disease entities, the nurse may not accurately interpret the patient's response. An example of this misunderstanding occurred repeatedly during validation studies when the nurses unquestioningly interpreted the patient's response "measles" as meaning rubeola, when the patient was referring to rubella.

Family history.

- Specific questions must be asked regarding the health of each immediate family member. The frequently asked "Do you have any children?" reveals nothing about the offsprings' health. Similarly, while investigating the chief complaint, the nurse may ask if anyone else in the family has the same problem. Again, this is an inadequate appraisal of the health of other family members, and is insufficient for credit.

- In addition to the specific familial diseases listed, nurses tend to enquire about such conditions as epilepsy, heart disease, and cancer. Any or all such questions support credit for the item: 'other major health problems with hereditary transmission.'

Functional inquiry.

- Note that there is no opportunity for credit to be given for a general query (or affirmative or negative response) for a system. For example, there is no score possible for a question such as: "Have you had any trouble with your eyes?"

- In response to a general question, however, the patient may elaborate on several specific items. Credit is then awarded for all items for which information has been elicited. The items 'photophobia,' 'discharge,' and 'vision' should be credited if, to the nurse's general question "Have you had any trouble with your eyes?" the patient replies "I can see O.K., but the light has been bothering me lately, and they're a bit runny."

Observer Evaluation: Standardizing Questions

The standardizing questions below are to be asked at the conclusion of the second history which the subject takes, in the absence of positive responses during the interviews, as indicated for each question.

In the event that the subject has received a positive response by one of the adult patients, she is credited for the other patient for the same items. In such circumstances, the applicable standardizing question is not to be asked.

a) If neither patient has a presenting complaint, nor any indication of abnormality (in the functional inquiry), ask:

"What additional questions would you have asked if this patient had come complaining of shortness of breath?"

b) Ask if you have not observed an interview of a patient with a previously established illness:

"If you had found that the patient had previously been diagnosed as having congestive heart failure, what else would you have asked him?"

c) If the subject has enquired about major illnesses, operations, or injuries and received negative responses, ask:

"If the patient had said that he had had infectious hepatitis, what else would you have wanted to know?"

d) If the nurse has enquired about diseases with familial predisposition and received negative responses, ask:

"If the patient had said that his mother had diabetes, would it have made any difference in your assessment of him?"

e) If the nurse has not interviewed a married patient, ask:

"What would you have asked if this patient had been married?"

f) If the nurse has enquired about the patients' families, and the patients have not had any children, ask:

"What would you have asked if the patient had had two children?"

Scoring

For all standardizing questions which have been asked, give credit on the observations of histories and the charting of them for those items which the nurse mentions in her responses.

Complete Routine Physical Examination

Individuals performing a physical examination will often use differing techniques because of personal preferences. The bank of objectives originally stipulated abnormal findings which CTN's must be capable of recognizing upon examination. However, during the development of the evaluation, the need became apparent for specification of the content of a routine complete physical examination in behavioral terms. The COMPLETE ROUTINE PHYSICAL EXAMINATION instrument was subsequently content validated during a workshop of the participating CTN universities in September, 1974. Specified criteria were therefore not imposed by the designer of the instrument, but by agreement of the content validators. While the evaluator's personal criteria for acceptable techniques may vary from those specified, he must assess according to the criteria of the instrument.

The following points provide additional background information regarding the instrument for observation evaluation of the physical examination.

a) The criteria are not totally inclusive as to detail. In the interest of brevity (essential to usability of the form), assumptions were made regarding the evaluator's knowledge of various techniques of examination. For example, it was assumed but not specified that during fundoscopy the ophthalmoscope should be held close to the examiner's forehead, with the index finger resting on the lens wheel and the nurse's forehead almost touching that of the patient. Reliability should be maintained by the consistent application of such background knowledge.

b) Some items which were validated as necessary component of a complete routine physical examination were not included in the instrument. Exclusion was necessitated by the difficulty in assessing whether or not the subject did complete the item, particularly in the absence of positive findings. Examples of items omitted are: 'listens to speech,' and 'smells breath.'

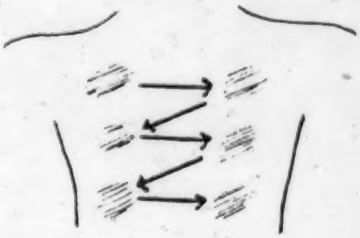
Observer Evaluation: Instructions

There are two columns applicable to each particular examination technique. The left column designates areas, functions and systems. The question implicit in the first column is, "Did she attempt to examine relevant to that particular item?" Place a check (☑) in the corresponding box if the nurse even attempts an examination related to that item, regardless of her degree of accuracy or correctness.

For almost all of the specific items in the left column, approved criteria are specified in the right column. To receive credit (also designated by placing a check appropriately), the nurse must perform the examination according to the technique spedified. The implicit question of the second column is, "Did her technique of examination meet the criteria?" For example, for the item 'palpates: fremitus,'

the schematic drawings below indicate appropriate scoring under various circumstances.

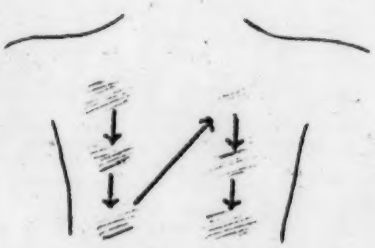
i)



If palpated front & back: ☑ ☑

If palpated back only: ☑ ☐

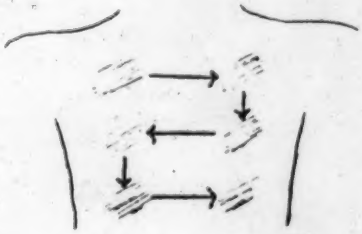
ii)



If palpated front & back: ☑ ☐

If palpated back only: ☑ ☐

iii)



If palpated front & back: ☑ ☑

If palpated back only: ☑ ☐

iv)

If patient speaks at all while nurse's hand rests on his back: ☑ ☐

All conditions of the criteria for acceptable performance (right-hand column) must be met for credit to be given. (For example, a common error during validation studies was the incorrect crediting of the item 'sharp/dull sensation' when the subjects tested only the anterior aspect of the limbs.) Because of this requirement, it is usual for more items in the left than the right column to be checked.

No partial marks may be given for any item in either column. In essence, the provision of the first column does allow a subject to obtain credit for attempting examination, although she may not be thoroughly competent.

No extra credit can be given for outstanding performance (e.g., the subject may do a more thorough musculoskeletal examination than required, but credit cannot be given for so doing). This again is a reflection of the theoretical framework of the evaluation which asks the question: "How many of the objectives do they meet?" (as opposed to the frequently encountered philosophy "How good are they?").

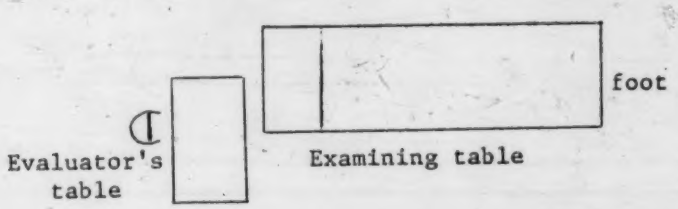
Some items on the form are easily overlooked (such as 'palpates pulses' which is credited if the radial pulse has been counted). Therefore, to ensure that rigorous consideration has been given to the scoring of all items, review the marking immediately after the completion of the physical examination.

If the subject indicates that she has been unable to perform a complete examination because of unavailability of equipment:

a) ask what equipment she would have used, and how she would have used it;
b) ask to see documentation of her request for such equipment;
c) credit those items which the nurse has adequately described IF the equipment has been requisitioned; and
d) list the missing equipment and date of requisitioning on the recorded physical sheet.

This measure will allow the subject to obtain credit for those items for which, through no fault of her own, she would otherwise not obtain credit.

As it is necessary to move rapidly from one page to another of the physical examination instrument, accurate scoring is facilitated by having the three pages spread out on a table. Place the table in a position which allows unobstructed observation of the majority of the examination while you are seated. During validation studies, the most appropriate arrangement of furniture in the room was found to be as illustrated below.



Specific Instructions

- Credit the following if the subject records them on the chart:

respiratory rate
pulse rhythm
pulse rate
apex rate

- Credit 'urinalysis' if the determination of all of the following are included: pH, sugar, acetone, blood, and albumin. Appropriate dipsticks or tablets may be used.

- The subject may refer to the patient's chart during the patient assessment, but referral to other reference sources is not allowed until the assessment has been completed.

Complete Paediatric History

The COMPLETE PAEDIATRIC HISTORY form is to be used for the assessment of the history taking for any patient less than sixteen years of age. The form is to be used in the same manner as that for the evaluation of adult histories (cf. pp. 126-127).

Standardizing Questions

The following standardizing questions are to be asked if you have not observed the taking of a paediatric history which would allow the subject to be scored on the items to which the questions pertain.

a) If the patient has not had a presenting complaint, ask:

"What else would you have asked if this child had been complaining of, or had signs of, abdominal pain?"

b) Ask, if the child did not have a previously established illness:

"If the mother had told you that her child was anemic, what else would you have asked her?"

c) If the nurse has asked about major illnesses, operations, or injuries and received negative responses, ask:

"If you had been told that this youngster had had 'a broken collarbone' what other questions would you have asked?"

d) If the subject has enquired about diseases with familial predisposition and received negative responses, ask:

"Would it have made any difference in your assessment of this child if you had found out that his brother was mentally retarded?"

e) If the nurse has enquired about the mother's pregnancy and there were no indications of a high risk pregnancy, ask:

"What difference would it have made if the mother had had any of the problems during her pregnancy which you asked about?"

f) If the subject receives a negative response to an enquiry regarding neonatal jaundice, ask:

"If the child had had jaundice as a newborn, would it have made any difference in your assessment of him?"

g) If you have not observed a history being taken of a child less than two years of age, ask:

"What other questions would you have asked the mother if the patient had been 18 months old?"

Scoring

For all standardizing questions asked, give credit on the observations of histories and the charting of them for those items which the subject mentions in her responses.

Suturing

It should be noted that the objectives from which this instrument was developed relate to a straight, clean laceration. The management of all such lacerations with which patients present during an on-site visit should be observed, as repeated scoring of the subject's suturing will lead to the best estimate of her ability.

Place a check (☑) in each box if the item is successfully completed according to the specified criteria. Several items will be done repeatedly during the suturing, but credit is given only if an item is properly done each and every time. If the nurse performs an item other than as specified, place a cross (☒) against that item.

When the subject has completed the suturing, complete the scoring by ensuring that all boxes have been marked.

Specific Instructions

- If bleeders are present and not ligated, place a cross in each box related to the achievement of haemostasis.
- If the nurse routinely sutures with a straight cutting needle, tell her that you would like to see her suture with a curved cutting needle. If she is unable to do so, give no credit for that item, or the item preceding it (re: needle holder).
- As zone regulations regarding tetanus protection differ, the appropriateness of that given is judged on those recommendations provided by the subject's zone.

Standardizing Questions

The standardizing questions are to be asked only if suturing has been observed of lacerations in which no bleeders were present.

a) "If bleeders had been present in the laceration(s) which you sutured, how would you have achieved haemostasis?"

b) If the subject mentions ligation, but not the type of suture she would use, ask:

"What kind and size of suture would you have used for ligatures?"

c) If ligatures have been discussed, ask:

"Where should the suture be cut following ligation?"

Scoring

On each sheet used to evaluate that subject's suturing of a laceration in which no bleeders were present, give credit for those items for which the nurse has given an accurate oral response.

INFORMATION FOR NURSES REGARDING CTN FIELD EVALUATION

As you have previously been informed, the evaluation has two main objectives:

i) to establish whether or not there is a difference in performance between nurses who have had the CTN course and those who have not; and

ii) to ascertain whether or not the CTN courses meet Medical Services Branch's need for ongoing education for nurses in nursing stations.

Consequently, both a group of CTN graduates and a control group of nurses who have not had the course have been randomly selected. The control group is crucial to the evaluation--without data regarding the skills and knowledge of nurses who have not taken the course, valid conclusions cannot be drawn about change in nursing practice following the CTN course. Additionally, the CTN group is divided into recent and non-recent graduates, to gain information regarding the influence of nursing station experience on performance following course completion.

While each university is responsible for student assessment upon course completion, it has not been shown that CTN's retain or practice their competencies after returning to their stations. To assess your nursing practice in the station, the evaluator will both be reviewing the charts of patients whom you have previously seen, and observing you with some of your patients during the time he is there.

The particular format of the evaluation in which you are taking part has been used in order to be as objective as possible. The evaluation is based on many specific criteria which the universities teaching the CTN courses have agreed a nurse should be able to meet upon course completion.

Additionally, the assessments are meant to be as standardized as possible. That is, all nurses must be asked to "run the same race." If the evaluation were based only on the patients who presented in the normal week, there could be vast differences in the types of patients seen in the various stations. Therefore, you are asked to do the following while the evaluator observes you with your patients:

a) a complete patient assessment, including history and physical of at least one male and one female adult patient;
b) a complete assessment of at least one obstetrical patient early in her pregnancy;
c) a complete history of at least one paediatric patient; and
d) suture a laceration.

The evaluator will review these patients' charts. Please arrange to do the patient assessments [a) to c)] during the first three days of the five day evaluator's visit, as there are also two short answer written examinations and one using audiovisuals which should be completed during the last two days.

It is re-emphasized that anonymity and confidentiality of results will be carefully preserved; no one nurse's score will be considered. As Medical Services Branch is interested in the effectiveness of the CTN course, the graduates' scores as a group will be compared to the scores for the group of nurses who have not taken such a course.

The written examinations can each be completed in approximately one-and-one-half hours, but with the exception of the simulation exam, there is no time limit for any portion of the evaluation. Specific instructions for the standardized tests follow.

Complete Patient Assessment

Assessment of patients is considered an extremely important aspect of nurses' responsibilities in an isolated community. While the evaluator is in your station, you are requested to do complete assessments of at least two patients (one male and one female) who are at least 16 years old. These assessments should be done during the first three days of the evaluator's visit..

The patients chosen for assessment should be freely ambulatory, and preferably conversant in English. If the latter is not the case, an interpreter must be present. In order to benefit the patients, those selected should be in need of an assessment either because their condition warrants such, or they have not had an examination for some time.

Explain to the patient that:

a) he will be having a thorough checkup;
b) the evaluator is there to see how nurses work, but not him as a patient; and
c) the evaluator, while being a doctor/nurse, will not be taking part in the examination, except as an observer.

The assessment should include a <u>complete history</u>, and a <u>complete routine physical examination</u>. To increase the thoroughness of your patient evaluation, proceed as though you had little previous knowledge regarding the patient. Do the history and physical as thoroughly as possible, following your own routine. Feel free to refer to the patient's chart. Recording of the findings may be done either during, or at the conclusion of the assessment.

The history is scored by the observer on the basis primarily of the information you elicit from the patient, and the thoroughness of the enquiries. The physical examination is scored on two things: (1) the thoroughness, and (2) the manner in which various techniques of examination are performed. There is no set order in which the assessment should be done.

When your charting is finished, the evaluator will check the recording of your findings, and may question you about certain aspects of history-taking.

Paediatric History

A complete history of at least one child must be taken during the evaluation, preferably during the first three days of the visit. The patient whom you select must be under 16 years of age, and preferably under school age. Unless you feel that the child can give his full history, a parent or guardian should be present, as should an interpreter if the family is not freely conversant in English.

No physical examination of the child is required as part of the field evaluation, although you may choose to complete the child's assessment as part of your care of him.

With the exception of those pertaining to physical examination, please follow the instructions presented on the previous page (under Complete Patient Assessment).

Written Examinations

The two written examinations will each take approximately one-and-one-half hours to do, and are on the subjects of:

a) Obstetrics and the Newborn; and
b) General and Paediatrics.

As these examinations are to be written following the evaluation of your patient assessments and paediatric history-taking, time should be set aside on the last two days of the visit for the written exams. The questions are all short answer, and include true or false, multiple choice, and 'fill in the blanks' types of format.

Scoring of the written examinations is on the basis of the number of correct answers.

Suturing

The one portion of the field evaluation which depends on the type of patient who presents during the evaluator's visit is the assessment of suturing. Should anyone come to the station with a laceration while the evaluator is there on his visit, please inform him of the patient's arrival so that he can assess your management of the patient.

Information should again be given to the patient as to the evaluator's role. Otherwise, no deviation from your usual care of such patients is necessary.

APPENDIX H

Nursing Stations Stratified by Community Size

Nursing Stations Stratified by Community Size

≥500	<500
Aklavik	Arctic Bay
Assumption	Belcher Island
Baker Lake	Broughton Island
Big Trout Lake	Chesterfield Inlet
Brochet	Clyde River
Cambridge Bay	Coral Harbour
Cape Dorset	Fort Franklin
Coppermine	Fort Good Hope
Cross Lake	Fort Liard
Eskimo Point	Fort Norman
Faro	Fort Resolution
Fort Chipewyan	Fort Wrigley
Fort Hope	Foxe (Hall Beach)
Fort McPherson	Fox Lake
Fort Providence	Gjoa Haven
Garden Hill	Grise Fjord
God's Lake Narrows	Holman Island
Great Whale River	Lake Harbour
Igloolik	Lansdowne House
Inoucdjouac	Old Crow
Kashechewan	Pelican Narrows
Little Grand Rapids	Pelly Bay
Nelson House	Repulse Bay
New Osnaburgh	Round Lake
Norman Wells	Shamattawa
Oxford House	Snowdrift
Paint Hills	Spence Bay
Pangnirtung	Sugluk
Pikangikun	Whale Cove
Pond Inlet	
Poplar River	
Povungnituk	
Pukatawagan	
Rankin Inlet	
Resolute Bay	
Romaine	
Rupert House	
St. Theresa Point	
Sandy Lake	
South Indian Lake	
Split Lake	
Tuktoyaktuk	
Watson Lake	